

us-09-668-119-1.rst

OM nucleic - nucleic search, using sw model

Title:	US-09-668-119-1
Perfect score:	1740
Sequence:	1 atgaggaagctgtgtgac.....cacattcgttcagccatgta 1740

• Searched: 11351937 segs, 5372889281 residues  
Total number of hits satisfying chosen parameters:

Database :

```

1: em_estfun:*
2: em_esthum:*
3: em_estlin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hlc:*
10: gb_est1:*
11: gb_est2:*
12: gb_hlc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

```

RESULT	14
AA664125/c	
LOCUS	
DEFINITION	AA664125 482 bp mRNA
ACCESSION	AC00604.s1 Stratiogene lung (#937210) Homo sapiens EST
VERSION	12-NOV-1997
KEYWORDS	IMAG:855654 3' similar to contains element MSRI MSRI repetitive element ; mRNA sequence.
SOURCE	AA664125
ORGANISM	AA664125.1 GI:2618116
	EST.
	human.
	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [estevenson.wustl.edu](mailto:estevenson.wustl.edu)  
This clone is available royalty-free through INL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information  
Seq primer: -40m3 fwd. ET from Amersham  
High quality sequence stop: 425.

FEATURES	Location/Qualifiers
source	1. .482

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="855654"
/clone_lib="Stratagene lung (#937210)"
/sex="male"
/dev_stage="72 years"
/lab_host="SOLR cells (kanamycin resistant)"
/notes="Organ: Lung; Vector: pBluescript SK-; Site.1: SCORL1; Site.2: XhoI; Cloned unidirectionally. Primer: Oligo dT, normal lung. Average insert size: 1.0 kb; UniZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor sequence: 5' CTCGAGATTTTTTTTTTTTTTTT 3'."
BASE COUNT
ORIGIN
51 a 113 c 176 g 142 t

```

[illegible]

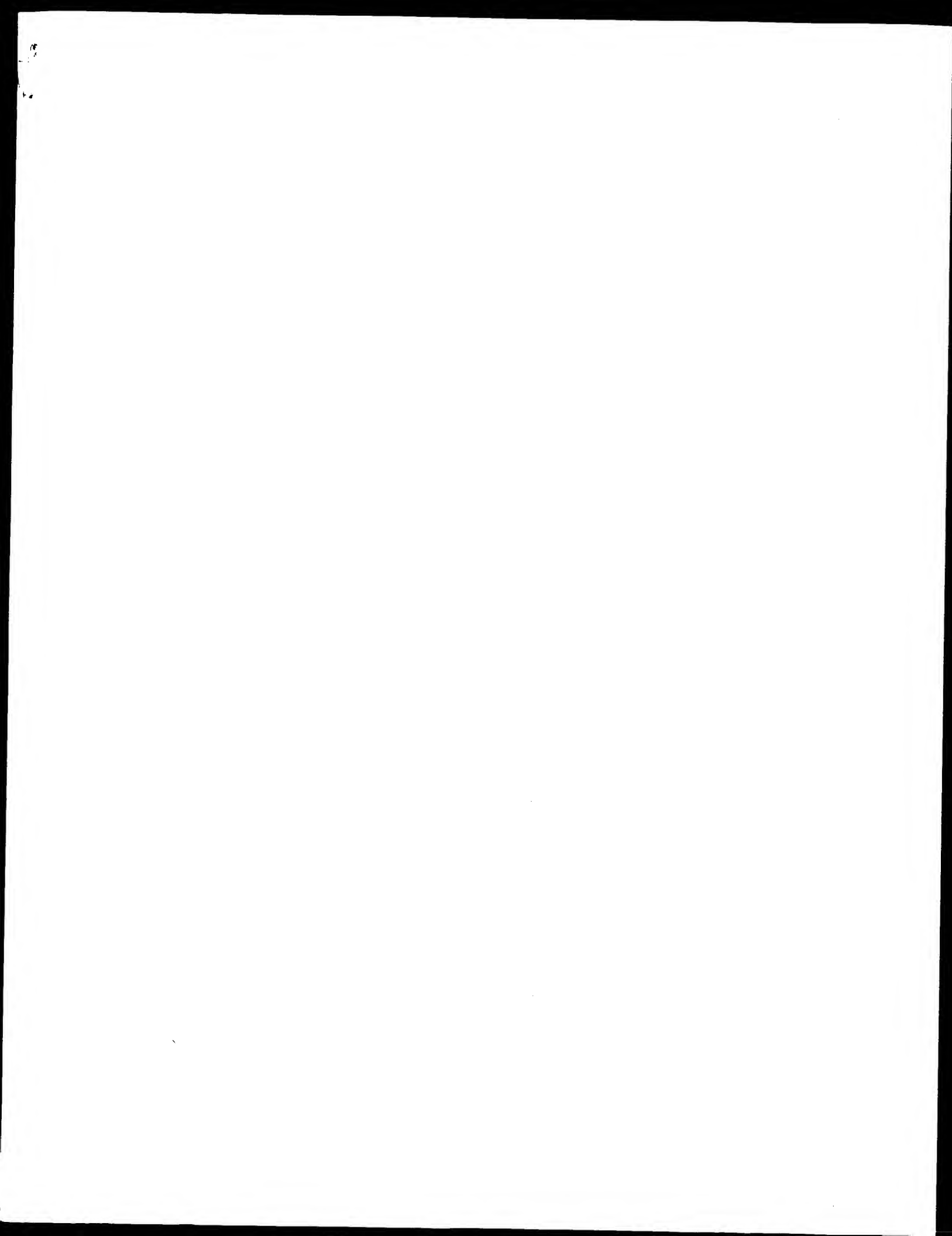






















us-09-668-119-1.rni

OM nucleic - nucleic search, using sw model

Title: US-09-668-119-1  
Perfect score: 1740  
Sequence: 1 atagagaaactgtgtgac.....cacattcgttcacagccatgta 1740

```

Searched:      351203 seqs, 113230333 residues
Total number of hits satisfying chosen parameters: 702406

```

```

Minimum DB seq length: 0
Maximum DB seq length: 20000000000
Post-processing: Minimum Match 0%

```

```
Maximum Match 100%
Listing first 45 summaries
```

```
Database : Issued Patients NA:*
1: /cgn2_6/prodata/2/1na/5A_CONH.seg:*
2: /cgn2_6/prodata/2/1na/5B_CONH.seg:*
3: /cgn2_6/prodata/2/1na/6A_CONH.seg:*
4: /cgn2_6/prodata/2/1na/6B_CONH.seg:*
5: /cgn2_6/prodata/2/1na/PCUTUS_CONH.seg:*
6: /cgn2_6/prodata/2/1na/backfiles.seg:
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	181.2	10.4	3489	2	US-08-728-323A-1	Sequence 1, Appl
2	181.2	10.4	3207	2	US-08-770-319-20	Sequence 20, Appl
3	181.2	10.4	3207	4	US-08-757-6692A-20	Sequence 20, Appl
4	171.8	9.9	543	6	5273901-6	Patent No. 5273901
5	170	9.8	533	6	54682709-5	Patent No. 54682709
6	142.6	8.2	397	3	US-09-253-601-3	Sequence 3, Appl
7	130.8	7.5	234	1	US-08-469-802B-3	Sequence 3, Appl
8	130.8	7.5	234	2	US-08-267-803B-3	Sequence 3, Appl
9	130	7.5	688	4	US-08-998-416-015	Sequence 915, Appl
10	126.6	7.3	6530	2	US-08-146-930-1	Sequence 1, Appl
11	126.6	7.3	6530	3	US-08-458-240-1	Sequence 1, Appl
12	126.6	7.3	6530	5	PCN-US93-03993-1	Sequence 1, Appl
13	120.6	6.9	477	4	US-09-135-994-1	Sequence 1, Appl
14	118.8	6.8	2580	3	US-09-050-863-2	Sequence 2, Appl
15	118.8	6.8	5452	2	US-09-130-114-1	Sequence 1, Appl
16	118.8	6.8	9600	4	US-08-910-647-1	Sequence 1, Appl
17	118.8	6.8	10596	1	US-07-884-811-15	Sequence 15, Appl
18	118.8	6.8	10596	1	US-07-885-971-15	Sequence 15, Appl
19	118.8	6.8	10596	1	US-08-087-783A-15	Sequence 15, Appl
20	118.8	6.8	10596	1	US-08-194-088B-15	Sequence 15, Appl
21	118.8	6.8	10596	2	US-08-194-087-15	Sequence 15, Appl
22	118.8	6.8	10596	5	PCN-US93-04648-15	Sequence 15, Appl
23	114.8	6.6	2793	1	US-08-209-747-1	Sequence 1, Appl
24	114.8	6.6	2793	1	US-08-458-298-1	Sequence 1, Appl
25	113.8	6.5	203	4	US-09-043-303-7	Sequence 7, Appl
26	113.4	6.5	2214	3	US-08-864-038A-1	Sequence 7, Appl
27	113.4	6.5	3331	3	US-08-864-038A-2	Sequence 2, Appl

RESULT 7  
US-08-469-802B-3  
; Sequence 3, Application US/08469802B  
; Patent No. 5741645  
GENERAL INFORMATION.

GENERAL INFORMATION:  
 APPLICANT: OIT, Harry T.  
 APPLICANT: Rarum, Laura P.W.  
 APPLICANT: Chung, Ming-Yi  
 APPLICANT: zozhbi, Huda Y.  
 TITLE OF INVENTION: Gene Sequence for Spino cerebellar Ataxia

Patent No. 5741645  
TITLE OF INVENTION: Type 1 and Method for Diagnosis  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Mueeling, Raasch, Gebhardt & Schwappach, P.A.  
STREET: 119 No. 5741645th Fourth Street, Suite 203

```

; CITY: Minneap
; STATE: MN
; COUNTRY: USA
; ZIP: 55401
; COMPANY: BARR

```

```

1  COMPUTER READABLE FORM:
2  MEDIUM TYPE: Floppy disk
3  COMPUTER: IBM PC compatible
4  OPERATING SYSTEM: PC-DOS/MS-DOS
5  SOFTWARE: PatentIn Release #1.0, Version #1.25
6  CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/469, 802E  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

NAME: Mueling, Ann M.  
REGISTRATION NUMBER: 33,977  
REFERENCE/DOCKET NUMBER: 110.00030101  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-305-1217

TELEPHONE: 012 303 1217  
TELEFAX: 612-305-1225  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 234 base pairs

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-469-802B-3

```

Query Match	7.58;	Score 130.8;	DB 1;	Length 234;
Best Local Similarity	75.74;			
Pred No	2	Re=20;		

```

Matches: 162; Conservative: 0; Mismatches: 52; Indels: 0; Gaps: 0;

```

370 cagaccacagctgcagctccagcaggtgcygctgcagcagcagcagcaacagcagcagctc 429

Db  
1 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 60

Oy 430 caacaccagcaccgacgcctatcaccaaccaaacccacgaacttc 489

[illegible][illegible]

490 cagugctcagcagagltgcacatgcagcagcagltccacagcagtaagltcagcagcagcagcagcag 549

[illegible]

QY 550 ctccagcagcagcagcagcagcatctaa 583

Db 181 CAGCAGCAGCAGCAGCAGCAGCAGCAGCACTCA 214

1











## alignment\_scores:

Quality: 2959.00 Length: 580  
 Ratio: 5.111 Gaps: 0  
 Percent Similarity: 99.828 Percent Identity: 99.483

## alignment\_block:

US-09-668-119-1 x AAM40299

Align seg 1/1 to: AAM40299 from: 1 to: 748

```

1  ATGAGAAAGCTGCTGGACACAGTAATCCAGCAAGATATGAGAG 50
27  MetArgLysAlaGlyValAlaHisSerLysSerLysAspMetGlu 43
51  CCAGTTTCTGGAAGCCAGACCCGGAGAGATACCTTTCTGCTGG 100
43  HisValPheLeuLysAlaLysThrArgAspGluTyrLeuSerLeuVal 60
101 CCAGGCTCATTTATCCATTTTCGAGACATTCATAACAAGAAATTCAGCT 150
60  LaArgLeuIleIleHisPheArgSpIleHisAsnLysSerGlnAla 76
151 TCCGTCAGTATCTCTATGATGACCTCCAGAGCTGACTGGCGACCTGC 200
77  SerValSerAspProMetAsnAlaLeuGlnSerLeuThrGlyProAl 93
201 TGCAGGAGCGCTGGAATGGCATGCTCCTCGGGGGCGCCGGACAGTCTC 250
93  aAlaGlyAlaAlaGlyIleGlyMetProProArgGlyProGlyGlnSer 110
251 TGGCGGGAGTGGTAGCTTTGGTGGCCATGGGACAGCCAAATGCTCTCTCA 300
110  euGlyGlyMetGlySerLeuGlyAlaMetGlyGlnProMetSerLeuSer 126
301 GGGAGGCGCTCTCTGGAGCTGGGGAGATGGCCCTCACAAGCTGGCTGT 350
127  GLyGlnProProProGlyThrSerGlyMetAlaProHisSerMetAlaVal 143
351 CGTGTCTACGCAACTCCACAGACCCAGCTGCAGTCCAGCAGGTGGCGG 400
143  ValSerThrAlaThrProGlnThrGlnLeuGlnLeuGlnValAlaLeu 160
401 TGGAGAGACAGCAGCAACAGCAGCACTTCACAGCAGCAGCGCGG 450
160  euGlnGlnGlnGlnGlnGlnGlnGlnPheGlnGlnGlnGlnAlaAla 176
451 CTACAGCAGCAGCAGCAGCAGCAGCAGCAGTTCAGCAGCAGCAGCAGC 500
177  LeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnAlaGlnGln 193
501 GAGTGCATGACAGCAGCAGCTCCAAAGCAGTAGTCAGCAGCAGCAGC 550
193  nSerAlaMetGlnGlnGlnPheGlnAlaValAlaGlnGlnGlnGlnGln 210
551 TCCAGCAGCAGCAGCAGCAGCAGCAGCAGCTTAATTAATTCATATCA 600
210  euGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 226
601 AATCAACAACAGATACAGCAGCAGCAGCAGCAGCTGCAGCGAATGAC 650
227  AsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 243
651 GCTGAGCTCCCAACAACAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 700
243  nLeuGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260
701 AGCAGCAGGCTTTGGAGGCCAGCAGCAGCAATTCAGCAGCAGCAGCAG 750
260  GlnGlnGlnAlaLeuGlnAlaGlnProProIleGlnGlnGlnGlnGln 276
751 CAGCCACAGCCTCCGCTCCAGGCTCTGCCAGCAGCAGCTGCAGCAGAT 800
277  GlnProGlnProProProSerGlnAlaLeuProGlnGlnLeuGlnGln 293
801  GCATCACACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 850
293  ThrHisThrGlnHisGlnHisGlnProProProGlnGlnGlnGlnPro 310
851  TTGCTCAGAACCAACCATCAACAATCCCGCAGAGTCCGACAGCCAGCT 900
310  AlaAlaGlnAsnGlnProSerGlnLeuProProGlnSerGlnThrGlnPro 326
901  TTGGTGTACAGGCGCAGCAGCTCTCCCTGACAAATGTTATACCAACC 950
327  LeuValSerGlnAlaGlnAlaLeuProGlyGlnMetLeuTyrThrGlnPro 343
951  ACCACTGAATTTGTCGAGCTCCGATGAGTGGGAGCAGCAGCCCGACAG 1000
343  OProLeuLysPheValArgAlaProMetValAlaGlnGlnProProValG 360
1001 AGCCCGAGTGCAGCAGCAGCAGCAGCAGTACAGACAGCTCAGCCTGCC 1050
360  InProGlnValGlnGlnGlnGlnThrAlaValGlnThrAlaGlnAlaAla 376
1051 CAGATGTGCTGCTCCCGGAGTCCAGGTCACAGCAGCAGCAGCAGCTCT 1100
377  GlnMetValAlaProGlyValGlnValSerGlnSerLeuProMetLeu 393
1101  GTTCCTGCGCGTCACCGGGCAGCAGCAGTGCAGACCCGCGAGTGCATGCC 1150
393  uSerSerProSerProGlyGlnGlnGlnValGlnThrProGlnSerMetPro 410
1151  CTCCCCCAGCGCTCCCGCGACGCCGCGCAGCCAGCTCAGACCCCAAC 1200
410  roProGlnProSerProGlnProGlyGlnProSerSerGlnProAsn 426
1201  TCCAACTCAGCTGCGGCTGCGCCCATCTCCAGTGGCTCCGCGCAG 1250
427  SerAsnValSerSerGlyProAlaProSerProSerSerPheLeuProSe 443
1251  CCCCTCACCGCAGCCCTCCAGAGCCCAAGTACGCGCGCGACCCACAGA 1300
443  rProSerProGlnProSerGlnSerProValThrAlaArgThrProGln 460
1301  ACTTCAGTGCCTCCCTCACTGAGCTTTAAACACACCTGTGACCCAGC 1350
460  snPheSerValProSerProGlyProLeuAsnThrProValAsnProSer 476
1351  TCTGTATGAGCCCAAGCTGGCTCCAGCCAGCTGAGAGCAGCAGTACCT 1400
477  SerValMetSerProAlaGlySerSerGlnAlaGlnGlnGlnGlnGln 493
1401  GGACAAGCTGAGAGCAGCTGTGCAAGTACATCGAGCCCTCGCCGCATGA 1450
493  uAspLysLeuLysGlnLeuSerLysTyrIleGlnProLeuArgArgMet 510
1451  TCAACAAGATCGACAGAAGACAGAGCAAGAAAGACCTGAGTAAGATG 1500
510  LeAsnLysIleAspLysAsnGlnAspArgLysLysAspLeuSerLysMet 526
1501  AAGAGCTTGTGACATTTGACAGACCCCTCAGACGGTGTCCCTCGAA 1550
527  LysSerLeuLeuAspIleLeuThrAspProSerLysArgCysProLeuLys 543
1551  GACCTTGCAAAAGTGGAGATGCGCTCGAGGAAGATCAAGATGACATGC 1599
543  sThrLeuGlnLysCysGlnIleAlaLeuGlnLysLeuLysAsnAspMet 560
1600  GGTGCCACTCCCGCAGCCGCGCTGCACCGCAGCAAGCAGCAGCAGTAC 1649
560  LaValProThrProProProProValProProThrLysGlnGlnTyr 576
1650  CTATGCCAGCGCGCTCTGATGCGCTGCGCACACATCGCTCACCTGT 1699

```



577 LeuCyG6InProLeuEuAspAlaValLeuAlaAsnIleArgSerProVa 593

1700 CTTCAACATTCCTGACCGACATTCGTCAGCCATG 1739

593 lPheAsnHisSerLeuTyrArgThrPheValProAlaMet 606

seq\_name: /SID52/gcgdata/geneseq/geneseq/AA2001.DAT:AA42085

seq\_documentation\_block:

ID AA42085 standard; Protein: 780 AA.

AC AA42085;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 7016.

Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokinetic; thrombolytic; drug screening; arthritis; inflammation; leukaemia.

OS Homo sapiens.

PN WO200153312-A1.

PD 26-JUL-2001.

PF 26-DEC-2000; 2000MO-US34263.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

PI Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;

PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

DR MPI: 2001-442253/47.

DR N-PSDB; AAT61241.

PT Novel nucleic acids and polypeptides, useful for treating disorders

PT such as central nervous system injuries -

PS Example 2; SEQ ID NO 7016; 10078bp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and

CC the encoded polypeptides (AA38642-AA42213) with nootropic,

CC immunosuppressant and cytostatic activity. The polynucleotides are useful

CC in gene therapy. A composition containing a polypeptide or polynucleotide

CC of the invention may be used to treat diseases of the peripheral nervous

CC system, such as peripheral nervous injuries, peripheral neuropathy and

alignment\_scores:  
Quality: 2860.00 Length: 581  
Ratio: 4.948 Gaps: 0  
Percent Similarity: 99.484 Percent Identity: 98.967

alignment\_block:  
US-09-668-119-1 x AA42085

Align seg 1/1 to: AA42085 from: 1 to: 780

```

1 ATGAGGAAAGCTGCTGTGGACACACAGTAATCCAGCAGATATGAGAG 50
|||||
58 MetArgLysAlaGlyValAlaHisSerLysSerSerLysAspMetGlu 74
|||||
74 rHisValPheLeuLysAlaLysThrArgAspIuTyrLeuSerLeuVal 91
|||||
91 laArgLeuLleIleHisPheArgAspIleHisAsnLysSerGlnAla 107
|||||
101 CCAGGCTCATTTTCATTTTGGAGCATTCATTAACAAGAAATTCAGCT 150
|||||
151 TCCGTCAGTATCTATGATGCACTCCAGACCTGAGCGGAGCTGC 200
|||||
108 SerValSerAspProMetAsnAlaLeuGlnSerLeuThrGlyProAl 124
|||||
201 TGGCGGAGCCGCTGGAATGGCATGCCCTCTCGGGGCCGAGCTGC 250
|||||
124 aAlaGlyAlaAlaGlyIleGlyMetProProArgGlyProGlyGlnSer 141
|||||
251 TGGCGGAGTGGGTAGCTTGGTCATGGGACGACCATGCTCTCA 300
|||||
141 euGlyIleMetGlySerLeuGlyAlaMetGlyGlnProMetSerLeuSer 157
|||||
301 GGGACACCCCTCTCTGGGACCTCGGGGATGGCCCTCACAGCATGGCTG 350
|||||
158 GlyGlnProProProGlyThrSerGlyMetAlaProHisSerMetAlaVa 174
|||||
351 CGGTCTACGGCACTCCACAGACCCAGCTGCTCCAGCAGGTGGC. G 399
|||||
174 lValSerThrAlaThrProGlnThrGlnLeuGlnLeuGlnValAla 191
|||||
400 CTCACACACAGCAGCAGCAGCAGCAGC. AGTTCCAGCAGCAGCAGCGG 448
|||||
191 laAlaAlaAlaAlaAlaThrAlaArgSerSerSerSerSerAlaArg 207
|||||
449 CGCTACAGCAGCAGCAGCAGCAGCAGCAGCAG. CAGTTCCAGGCTCA 497
|||||
208 ArgTyrSerSerSerSerSerSerSerSerSerSerSerSerAlaGln 224
|||||
498 GCAGAGTCCCATGACACAGCAGCTTCCACAGCAGTATGTCAGCAGCAGC 547
|||||
224 ngInSerAlaMetGlnGlnGlnPheGlnAlaValAlaGlnGlnGlnGln 241
|||||
548 AGTTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTAAATTCATCA 597
|||||
241 lnuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 257
|||||
598 CAAATTCAGCAGCAGTATGAGCAGCAGCAGCAGCAGCAGCTCAGGATAC 647
|||||
258 GlnAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 274
|||||
648 ACAGCTGACGCTCCACACACAGCAGCAGCAGCAGCAGCAGCAGCAGC 697
|||||
274 agInLeuGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 291
|||||
698 AGCAGCAGCAGGCTTGGAGCCAGCAGCAGCAGCAGTATGAGCAGCAGGAT 747
|||||
291 lngInGlnGlnAlaLeuGlnAlaGlnAlaGlnProProIleGlnGlnProMet 307
|||||
748 CAGCAGCCACAGCCTCCGCTCCAGGCTGCTGCCAGCAGCTGCAGCAGCA 797

```











CC predicting, diagnosing, grading, staging, monitoring and prognosing  
 CC diseases of the human breast, particularly those diseases with polygenic  
 CC aetiology. The diseases include: breast cancer, disorders of development,  
 CC inflammatory diseases of the breast, fibrocystic changes, proliferative  
 CC breast disease and non-carcinoma tumours.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 79 AA;

alignment\_scores:  
 Quality: 390.00 Length: 79  
 Ratio: 4.937 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-09-668-119-1 x AAM01992 ..

Align seq 1/1 to: AAM01992 from: 1 to: 79

```

376 CAGCTGCAGTCCAGCGGTGGCTCCAGCAGCAGCAGCAGCA 425
      |||||||
1  GlnLeuInLeuGlnGlnValAlaLeuGlnGlnGlnGlnGln 17
426 GTTCACAGCAGCAGCAGCGCGCTCAAGCAGCAGCAGCAGC 475
      |||||||
17 nphGlnGlnGlnGlnGlnAlaLeuGlnGlnGlnGlnGlnGln 34
476 AACAGCAGCGTTCAGGCTCAGCAGAGTCCATGCAGCAGCA 525
      |||||||
34 InGlnGlnGlnGlnGlnAlaGlnGlnGlnGlnGlnGlnGln 50
526 GCAGTAGTCAGCAGCAGCAGCAGTCCAGCAGCAGCAGCAGCA 575
      |||||||
51 AlaValAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 67
576 GCATCTAATTAAATGCATCATCAAAATCAGCAACAG 612
      |||||||
67 nHISLeuIleLysLeuHISHisGlnAsnGlnGlnGln 79

```

seq\_name: /SID52/gcdata/geneseq/geneseq/AA198.DAT:AAW37856  
 seq\_documentation\_block:  
 ID AAW37856 standard; Protein; 1004 AA.  
 AC  
 XX AAW37856;

```

DT 28-AUG-1998 (first entry)
XX
DE Human polyhomeotic 1 (hph1) protein.
XX
KW Polyhomeotic 1; hph1 gene; oncogene; human; neoplasia;  

  KW dysplasia; hyperplasia; diagnosis; tumour; metastasis; therapy.
XX
OS Homo sapiens.
XX
FN WO9807860-A1.
XX
PD 26-FEB-1998.
XX
PF 22-AUG-1997; 97WO-US14886.
XX
PR 06-FEB-1997; 97US-0036939.  

  PR 23-AUG-1996; 96US-0024349.  

  PR 04-DEC-1996; 96US-0031569.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Randazzo F;
XX
DR WPI; 1998-169164/15.

```

DR N-PSDB: AAV29009.

XX  
 XX Isolated human polyhomeotic 1 oncogene - used to develop products  
 PR for diagnosis and therapy of proliferative and developmental  
 PR disorders, e.g. neoplasia, dysplasia or hyperplasia  
 XX  
 XX

PS Claim 1: Page 28-30; 39pp: English.

XX This polypeptide comprises a novel human polyhomeotic 1 protein  
 CC encoded by the hph1 oncogene that is implicated in neoplastic  
 CC disorders. Also claimed are: (1) an hph1 polypeptide consisting of  
 CC at least 22 contiguous amino acids; (2) an hph1 fusion protein; (3)  
 CC antibodies which specifically bind to a hph1 protein; (4) an  
 CC expression construct for expressing all or a portion of hph1  
 CC protein; (5) a homologously recombinant cell; (6) a method of  
 CC identifying neoplastic tissue of a human, where over-expression of  
 CC the hph1 gene identifies the tissue as being neoplastic; (7) a  
 CC method to aid in the diagnosis or prognosis of neoplasia, where a  
 CC difference in the hph1 gene, mRNA, or protein between a first and  
 CC second tissues indicates neoplasia in the first tissue; (8) a method  
 CC to aid in detecting a genetic predisposition to neoplasia; (9) a  
 CC method of identifying a human chromosome 12; and (10) a method of  
 CC inducing a cell to de-differentiate by contact with a hph1 gene or  
 CC expression product. The products can also be used to inhibit hph1  
 CC growth. They can be used to treat e.g. tumours, anhydric hereditary  
 CC ecodermal dysplasia, congenital alveolar dysplasia, epithelial  
 CC dysplasia of the cervix, fibrous dysplasia of bone, or mammary  
 CC dysplasia, endometrial, adrenal, breast, prostate, or thyroid  
 CC hyperplasia or pseudo-epitheliomatous hyperplasia of the skin.  
 CC Even in disorders in which hph1 mutations are not implicated,  
 CC down-regulation or inhibition of hph1 expression can have therapeutic  
 CC application. In these disorders, decreasing hph1 expression can help  
 CC to suppress tumours. Similarly, in tumours where hph1 expression is  
 CC not aberrant, effecting hph1 down-regulation can suppress metastases.  
 XX  
 XX Sequence 1004 AA;

alignment\_scores:  
 Quality: 389.00 Length: 522  
 Ratio: 1.379 Gaps: 20  
 Percent Similarity: 54.023 Percent Identity: 28.544

alignment\_block:  
 US-09-668-119-1 x AAW37856 ..

Align seq 1/1 to: AAW37856 from: 1 to: 1004

```

178 CAGAGCCTGACTGGCGGAGCTGCTGGGAGCGCGTGGATTCGATGCC 227
      |||||||
260 GlnSerLeuAsnLeuSerGlnAlaGlyLysGlySerGlyAsnSerIlePr 276
228 T.....CCTGGGGGCCCGGAGCAGTCTCGGCGGATGGATGA 265
      ||| ||| |||||
276 OGlySerMetGlyProGlyGlyGlyGlnAlaHisGlyLysGlyLeuGly 293
266 GCTTGGTGGCATGGGACAGCAATGTCTCTCAGGAGCGCGCTCCT 315
      : : : : :
293 InLeuProSerSerGly.....MetGlyGlyLysSerGlyProAlaGly 307
316 GGGACCTCGGGGATGGCCCTCAGCAGATGGCTGTGCTCAGCGCAAC 365
      ||||| : : : ||| : : : ||| : : : ||| : : : |||
308 G1yThrGlyAlaValGlnProLeuProAlaAlaGlnThrValThrValSe 324
366 TCACAGACCCAGCTCAGCTCAGCAGAGTGGCGCTCAGCAGCAGCAGC 415
      : : : : : : : : : : : : : : : : : : : : : : :
324 rgInGlySerGlnThrGlnAlaGlnSerAlaAlaAlaLysLysAlaGlnA 341
416 AACAGCAGCGTTCAGCAGCAGCAGCAGCGCGCTCAGCAGCAGCAG 465
      : : : : : ||||| : : : ||| : : :
341 IaSpGlySerGlyGlnGlnAsnValGlyMetAsnLeuThrArgThrAla 357

```







CC 12p13, a region which is frequently lost in non-small cell lung cancer  
 CC and breast cancer. A method of identifying neoplastic tissue of a  
 CC human comprising comparing the expression of a hpl1 gene in a tissue of  
 CC a human suspected of being neoplastic with the expression of a hpl1 gene  
 CC in a tissue of the human which is normal. Under-expression of the hpl1  
 CC gene identifies the subject as having neoplastic tissue. The hpl1  
 CC oncogene functions to suppress neoplasia and dysplastic or hyperplastic  
 CC cell growth as well as to induce cellular differentiation. The cDNA,  
 CC protein and vectors can be used as diagnostic and therapeutic tools for  
 CC of a human chromosome 12. They can be used for the detection, diagnosis  
 CC or prognosis of neoplasia or for detecting a genetic predisposition to  
 CC neoplasia. They can also be used to treat tumours.

XX Sequence 1004 AA;

# alignment\_scores:

Quality: 389.00 Length: 522  
 Ratio: 1.379 Gaps: 20  
 Percent Similarity: 54.023 Percent Identity: 28.544

## alignment block:

us-09-668-119-1 x AAW52830 ..

Align seg 1/1 to: AAW52830 from: 1 to: 1004

```

178 CAGAGCCCTGCTGGGAGCCCTGGCGGAGCCCGCTGGAATGGCATGCC 227
    |||||||: : : : |||: |||: : : : |||: : : : |||:
260 GlnserLeuasnleuSerGlnAlaGlyGlySerGlyAsnSerIlePr 276
    : : : : : : : : : : : : : : : : : : : : : : : :
228 T.....CCTCGGGGCCCCGGGAGCATCTCTGGCGGATGGGTA 265
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
276 OGlySerMetGlyProGlyGlyGlyGlyGlnAlaHisGlyGlyLeuGly 293
    : : : : : : : : : : : : : : : : : : : : : : : :
266 GCTTGGTGGCATGGGACAGCCATGCTCTCGAGGGACGGCCCTCT 315
    : : : : : : : : : : : : : : : : : : : : : : : :
293 InleuProSerSerGly.....MetGlyGlySerGlyProAlaGly 307
    : : : : : : : : : : : : : : : : : : : : : : : :
316 GGGACCTCGGGGATGGCCCTCACAGCATGCTGCTGCTGCGGAC 365
    |||||||: : : : ||| ||| ||| ||| ||| ||| ||| |||
308 GLyThrGlyAlaValAlaGlnProLeuProAlaAlaGlnThrValTh 324
    : : : : : : : : : : : : : : : : : : : : : : : :
366 TCACAGACCCAGCTGACGCTCCAGAGGTGGCGCTGCAGACAGCAG 415
    : : : : : : : : : : : : : : : : : : : : : : : :
324 rGlnGlySerGlnThrGlnAlaGlnSerAlaAlaAlaGlyAlaGln 341
    : : : : : : : : : : : : : : : : : : : : : : : :
416 AACAGGAGCAGTTCACAGCAGCAGCAGCGCGCTACAGCAGCAGCAG 465
    : : : : : : : : : : : : : : : : : : : : : : : :
341 lAsnArgGlySerGlyGlnGlnAsnValGlyMetAsnLeuThrArgTh 357
    : : : : : : : : : : : : : : : : : : : : : : : :
466 CAGCAGCAGCAGCAGCAGCAGTTCAGGCTCAGCAGAGTGCATGCAGCA 515
    : : : : : : : : : : : : : : : : : : : : : : : :
358 ThrProAlaProSerGlnThrLeuIleSerSerAlaThrThrThrGln 374
    : : : : : : : : : : : : : : : : : : : : : : : :
516 GCGGTTCACAGCAGTGTGTCAGCAGCAGCAGCAGCAGCAGCAGCAG 565
    ||| : : : : : : : : : : : : : : : : : : : : : :
374 eGlnProHisSerLeuIleGlnGlnGlnGlnGlnGlnGlnGlnGln 391
    : : : : : : : : : : : : : : : : : : : : : : : :
566 AGCAGCAG.....CAGCATCTAATTAATTCATCATCAAAATGAGCA 609
    : : : : : : : : : : : : : : : : : : : : : : : :
391 yselValValIleGlnGlnGlnIleAlaIleHis.....GlnGln 405
    : : : : : : : : : : : : : : : : : : : : : : : :
610 CAGATCAGCAGCAGCAGCAGCAGTTCAGCAGATGCA...CAGCTGCA 656
    ||| : : : : : : : : : : : : : : : : : : : : : :
406 GlnPheGlnHisArgGlnSerGlnLeuHisThrAlaThrHisLeuGln 422
    : : : : : : : : : : : : : : : : : : : : : : : :
657 GCTCCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 706
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
422 nleuAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 439
    : : : : : : : : : : : : : : : : : : : : : : : :
707 AG.....GCTTGGAGCCCGCAGCCCAATTCAGCAGCAGCAGCAG 744
    || : : : : : : : : : : : : : : : : : : : : : :

```

```

439 lnrGlnAlaThrThrLeuThrAlaProGlnProGlnValProPro 455
    : : : : : : : : : : : : : : : : : : : : : : : :
745 ATGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 794
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
456 ThrGln.....GlnValProProSerGlnSer.....Gln 465
    : : : : : : : : : : : : : : : : : : : : : : : :
795 GCGATGATGATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 844
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
465 nGlnGlnAlaGlnThrLeuValValGln..... 474
    : : : : : : : : : : : : : : : : : : : : : : : :
845 CTCACATCTGCTCAGACCAACCATGACATGCGCCGACAGTCGAGCAG 894
    ||| : : : : : : : : : : : : : : : : : : : : : :
475 ..ProMetLeuGlnSerSerProLeuSerLeuProProAlaAlaPro 490
    : : : : : : : : : : : : : : : : : : : : : : : :
895 CAGCCTTTGGTGTACAGAGCGGAGAGTCTCCCTGGACAAATGTGTAA 944
    : : : : : : : : : : : : : : : : : : : : : : : :
491 LysProProIlePro..... 981
    : : : : : : : : : : : : : : : : : : : : : : : :
945 CCAACCCAGCAGTGAATTTGTCGAGCTCCGATGGTG..... 981
    : : : : : : : : : : : : : : : : : : : : : : : :
498 rLysProProValAlaProIleLysProProGlnLeuGlyAlaAlaLys 515
    : : : : : : : : : : : : : : : : : : : : : : : :
982 .....GTGAGCAGCCCCCAGTCGAGCCAGCCAGTCGAGCAGCAGCAG 1023
    : : : : : : : : : : : : : : : : : : : : : : : :
515 eSerAlaAlaGlnGlnProPro.....ProHisIleProValGlnVal 529
    : : : : : : : : : : : : : : : : : : : : : : : :
1024 ACAGCAGTACAGACAGCTCAGGCTCCGACAGTGTGCTCCGAGTCA 1073
    : : : : : : : : : : : : : : : : : : : : : : : :
530 ValGlyThrArgGlnProGlyThrAlaGlnAlaGlnAlaLeuGlyLeu 546
    : : : : : : : : : : : : : : : : : : : : : : : :
1074 GGTACAGCCAGCAGCAGCAGCAGCAGTCCGATGCTGCTCGCGGCGG 1123
    : : : : : : : : : : : : : : : : : : : : : : : :
546 aGlnLeuAlaAlaAlaValPro..... 553
    : : : : : : : : : : : : : : : : : : : : : : : :
1124 AGTGCAGACCCCGCAGTGCATGACCCCTCCCGCCAGCCGTCGCCAG 1173
    ||| : : : : : : : : : : : : : : : : : : : : : :
554 .....ThrSerArgGlyMetProGlyThrValGln...SerGlyGln 566
    : : : : : : : : : : : : : : : : : : : : : : : :
1174 CCGGCGCAGCCAGCTCAGACGCCCACTGCAAGTCAAGCTGTGAGCC 1223
    : : : : : : : : : : : : : : : : : : : : : : : :
567 AlaHisLeuAlaSerSerProProSerSerGlnAlaProGlyAlaLeu 583
    : : : : : : : : : : : : : : : : : : : : : : : :
1224 CCCATCTCCCACTGCTGCTCCAGGCCCTCCAGCCGCGCCAGC 1272
    ||| : : : : : : : : : : : : : : : : : : : : : :
588 nGlyLysProProThrThrLeuAlaProGlyMetThrLeuAlaProVal 600
    : : : : : : : : : : : : : : : : : : : : : : : :
1273 .....AGCCAGTGCAGC 1284
    : : : : : : : : : : : : : : : : : : : : : : : :
600 lYThrAlaHisValValValLysGlyGlyAlaThrThrHisSerProVal 616
    : : : : : : : : : : : : : : : : : : : : : : : :
1285 GCGCGGAGCCCGCAGACCTGCAGTGCCTCCAGCTGACCTTTAAAC 1334
    ||| : : : : : : : : : : : : : : : : : : : : : :
617 AlaGlnValProAlaAlaPheGlyMetGlnSer.....ValHisLe 630
    : : : : : : : : : : : : : : : : : : : : : : : :
1335 ACCTGTGAACCCAGCTGTGTCATGAGCCAGCTGAGCTCCAGCAGCAG 1384
    ||| : : : : : : : : : : : : : : : : : : : : : :
630 uProGlyLysProGlnThrLeu.....AlaValLysArgLysAla 644
    : : : : : : : : : : : : : : : : : : : : : : : :
1385 AGGAGCAGCAGTACCTGAGCAGCAGTGAAGCAGCTGTGAATGCATG 1434
    : : : : : : : : : : : : : : : : : : : : : : : :
644 spSerGlnGlnGlnArgAspArgValSerThrLeuGlySerMetLeuPro 660
    : : : : : : : : : : : : : : : : : : : : : : : :
1435 CCGCTGCGCGCATGATCAGACAGATGACAGACAGACAGAGAAATA 1484
    : : : : : : : : : : : : : : : : : : : : : : : :
661 AlaLysAlaSerProValAlaGlnSerProLysValMetAspGlyLys 677
    : : : : : : : : : : : : : : : : : : : : : : : :
1485 GGACCTG...AGTAAGATGAAGACCTTCTGACAT..... 1518
    : : : : : : : : : : : : : : : : : : : : : : : :
677 rSerLeuGlyLysLysAlaGlnSerValAlaAsnValAlaAsnThr 694
    : : : : : : : : : : : : : : : : : : : : : : : :
1519 .....CTGACAGACCCCTCGAAGCGGTGCCCTG 1548
    : : : : : : : : : : : : : : : : : : : : : : : :
694 roSerSerGlyLeuValAlaLeuThr..ProAlaProSerValProPro 710
    : : : : : : : : : : : : : : : : : : : : : : : :

```















```

469 .....CAGCAGCAACAGCAGCTTC..... 489
1193 uThrSerProValProGlyGlnGlnGlnGlnPheIleAsnAla 1210
490 .....CAGGCTCAGCAGAGTGCATCCAGCAGCAGCTTC 522
1210 snGlyGlyThrGlyAlaAsnProGlnIleuSerGluIleMetLysGln... 1225
523 CAGCAGTAGTCAGCAGCAGCAGCAGCTCCAGCAGCAGCAGCAGCA 572
1226 ArgHisIleHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1242
573 GCAGCATCTAATT.....AAATTGCATC 595
1242 nGlyMetLeuIleuProGlnSerProPheSerAsnSerThrProLeuGln 1259
596 ATCAAAATCAGCAGACAGTACAGCAGCAGCAGCAGCAGCAGCAGCA 645
1259 InGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnAlaIleThrSerAsn 1275
646 GCACAGCTGAGCTCCAAACACAGCAGCAGCAGCAGCAGCAGCA 695
1276 SerPheSerSerPrometGlnGlnGlnGlnGlnGlnGlnGlnGln 1292
696 GCAGCAGCA..... 704
1292 nGlnLysProGlySerValLeuAsnAsnMetProProThrProThrSerL 1309
705 .....GCAGGCTTTGGAGCCGAGCCGAGCCGAGCCGAGCCGAGCC 740
1309 euGlnAlaIleuAsnAlaGlyAlaGlyAlaProGlyThrGlyLysSerAla 1325
741 ACCGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 778
1326 SerAsnValIleThrValSerAlaProSerProSerProGlyPheLeuSerAs 1342
779 TGCCCGCAGCAGCTGCAGCAGATGATCAGCAGCAGCAGCAGCAGCCGCA 828
1342 nGlyProSerIleGlyThrProSerAsnAsnAsnAsnSerSerSerAlaA 1359
829 CCACAGCCGCA.....GCAGCCTTCAGTGGTCAGACCAACCATCACA 872
1359 snAsnAsnProProSerValSerSerLeuMetGlnGlnProLeuSerAsn 1375
873 ACTGCCGCCACA.....GTGCAGAGCC 895
1376 ArgProGlyThrProProTyrIleProAlaSerProValProAlaThrSe 1392
896 AGCCTTGGTGTGCAGAGCGCAGAGCTCCCTGGACAATGTTGTAATACC 945
1392 rAlaSerGlyLeuAlaIleAlaSerSerThrPro.....A 1403
946 CAACCCAGCTGAATTTGTCAGAGCTCGAGTGGTGAGCAGCAGCCGCC 995
1403 laseRAlaAlaIleAlaIleThrCysAlaSerSerGlySerIleSerAsnSerSer 1419
996 AGTGACCCGAGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1045
1420 SerGlyAlaIleThrAlaIleGlyAlaSerSerThrSerSerSerSer.. 1435
1046 CTGCCAGATGTGGTCCCGAGTCA...CGTCAGCAGCAGCAGCAGCAGC 1092
1436 .....AlaGlySerGlyThrProLeuSerSerValSerThrProT 1449
1093 CCATATCTGCTCCGCGCTCAGCCGAGCAGCAGCAGCAGCAGCAGCAGCAGC 1142
1449 hSerIleAlaIleThrAlaIleThrSerSerGlyGlyGlyGlyGlyGlyGly 1465
1143 GATGCC...CCCTCCCGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1189
1466 AsnAlaGlyGlyGlySerSerThrThrProAlaSerAsnProLeuLeu 1482

```

```

1190 CACAGCCCACTCCACACTCAGCTTGGC..... 1218
1482 uMetSer.GlyGlyThrAlaIleGlyGlyThrGlyAlaIleThrThrThr 1498
1219 CCTGCCCATCTCCAGTAGTCTCTCCGAGCCGAGCCGAGCAGCAGCAGCAGC 1268
1499 SerThrSerSerSerSerArgMetMetSerSerSerSerSerSerSerSer 1515
1269 CCAGAGCCCACTGAGCGGCGGAGCCGAGCAGAACTTCAGTTCCTCCAC 1318
1515 rGlMetAlaIleAlaLeuGlnAlaAlaIleAlaArgAspAsnAspAspGluThr 1532
1319 CTGAGCCTTTAAACACACTGTGACACCCAGCTCTGTGATGAGCCAGCT 1368
1532 rSerProSerGlyGlyAsnThrAsnIleSerGlyGlySerGlyAlaAla 1548
1369 GGCCTCAGCCAGGCTGAGAGCAGCAGTACCTGAGCAGCAGCAGCAGCAG 1416
1549 GlyIleMetAlaSerLysGlyLys.....LeuAspSerIleLysGln 1562

seq_name: /STD2/gcgdata/geneseq/geneseq/AA2000.DAT:AAV74791
seq_documentation_block:
ID   AAV74791 standard; Protein; 233 AA.
XX
AC   AAV74791;
XX
DT   21-MAR-2000 (first entry)
XX
DE   Neisseria meningitidis ORF 263 protein sequence SEQ ID NO:1056.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   W09957280-A2.
XX
PD   11-NOV-1999.
XX
PE   30-APR-1999; 99WO-US09346.
XX
PR   01-MAY-1998; 98US-0083758.
PR   31-JUL-1998; 98US-0094869.
PR   02-SEP-1998; 98US-0098994.
PR   02-SEP-1998; 98US-0099062.
PR   09-OCT-1998; 98US-0103749.
PR   09-OCT-1998; 98US-0103796.
PR   09-OCT-1998; 98US-0103796.
PR   25-FEB-1999; 99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI   Petersen J, Piza M, Rappunoli R, Ratti G, Scalato E, Scarselli M;
PI   Tetteijn H, Venter JC.
XX
DR   WPI: 2000-062150/05.
DR   N-PSDB; AA253553.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics
XX
PS   Claim 2; Page 606; 1453pp; English.
XX
CC   AA253015 to AA254536, AA254577 to AA254615, and AAV74253 to AAV75941
CC   represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC   and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as

```







CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

XX  
 SO Sequence 307 AA:

alignment scores:

Quality: 336.50 Length: 317  
 Ratio: 2.588 Gaps: 19  
 Percent Similarity: 41.009 Percent Identity: 27.445

alignment block:

US-09-668-119-1/rev x AAB95124 ..

Align seg 1/1 to: AAB95124 from: 1 to: 307

```

1133 GTCTGCACCTGCTGCGCGGTGACGGGAGAGACATGCGGAGAGCTGCT 1084
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
11 ILeCysValCysValPro.....ValCysValGlyAlaHisLe 23
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1083 CTGGCTGACCTGGACTCCGGAGACACCATCTGCGGAGCTGAGCTGCT 1034
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
23 U.....CysValTrrValHisLeCysValC 32
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1033 GTACTGCTGCTGC.....TGCCTGCTGACCTGCTG 1005
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
32 ysValAlaGLeuCysGlyCysThrPheValCysValCysAlaCysValCys 48
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1004 GGCCTGACCTGGGGGCTGCTGACACCATCGAGCTGGCAATTTTCAG 955
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
49 AlAcysValCysGlyCysLeuCysVal..... 57
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
954 TGGGTGGTGGATATACAACTTTGTCCAGGAGAGCTTGGCCCTGTGACA 905
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
58 .....TrrGlyThrPheValCys.....ValCysAlaCys.... 67
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
904 CCAAGAGCTGAGTC.....TGCACCTGTCGGGAGCTGTGATGCTTGG 861
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
68 .....ValTrrValHisLeCysValCysAlaSerValCys..... 79
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
860 TTCGACGACACTGAGAGCTGCTGGGGCTGTGCGGCTGCTGCTGCTG 811
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
80 .....ValProValCysGlyAlaHisLeuCysVal 90
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
810 TGTGTGATGC.....ATCTGC...TGCAGCTGCTGGGCGAGAG 776
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
90 IArrLeuCysValGlyAlaHisLeCysValCysAlaCys..... 103
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
775 CCTGGAGAGCGGAGGCTGTGCTGC.....TGC 747
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
104 .....ValCysGlyCysThrPheValCysValCysAlaCys 115
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
746 ATCGGTGGCTGC...TGAATTGGGCTGGGCTCCCAAGCTGCTGCTG 700
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
116 ValCysGlyCysLeuCysValCysGlyAlaHisLeuCysValCysValCy 132
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
699 CTGCTGC.....TCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 665
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
132 sleuCysValGlyAlaHisLeuCysValCysLeuCysValCysAlaCysV 149
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
664 GTTGGAGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 630
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
149 alTrrGlyCysThrPheValCysValCysAlaCysValTrrGlyCysThr 165
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
629 .....TGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 569
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
166 PheValCysValCysAlaCysValCysGlyCysThrPhe..... 178
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
588 TTTATTAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 539
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
179 .ValCysValCysAlaCysValCysGlyCysThrPheValCysValCysL 195
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
538 GCTGCACTACTGCTTGGACTGCTGCTGCAATGCACTGCTGCTGAGCCTGG 489
  
```

```

|||||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
195 euCysValTrrValHisLeCysValCysValCysLeuCysValTrrVal 211
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
488 AACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 454
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
212 HisLeCysValCysAlaCysValTrrValHisLeCysValCysValPr 228
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
453 TAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 404
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
228 oValCysValCysLeuCysValTrrValHisLeCysValCysValCysL 245
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
403 GCAGCGCACCTGCTGAGACTGAGCTGCTGCTGCTGCTGCTGCTGAC 354
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
245 euCysValTrrValHisLeCysValCysValCysAlaCysValCys... 260
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
353 AGCAGACCATGCTGTGAGGGCCATCCCGAGAGTCCAGAGGAGGCTG 304
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
261 .....ValAlaGlyProGlyCys 266
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
303 C 303
      |||:||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
266 s 266
  
```

seq.name: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAV81609

seq\_documentation\_block:

ID AAV81609 standard; Protein: 1237 AA.

AAV81609:

24-MAY-2000 (first entry)

Streptococcus pneumoniae type 4 protein sequence #109.

Streptococcus pneumoniae; vaccine; screening; protein antigen;  
 antibacterial; antiinflammatory; meningitis; infection; diagnosis;  
 pneumococcal disease.

Streptococcus pneumoniae.

W0200006737-A2.

10-FEB-2000.

27-JUL-1999; 99WO-GB02451.

27-JUL-1998; 98GB-0016337.

19-MAR-1999; 99US-0125164.

(MICR-) MICROBIAL TECHNIQS LTD.

Gilbert CFG, Hansdro PM;

WPI: 2000-195300/17.

New Streptococcal protein, useful as a vaccine, for diagnosis of  
 PT pneumococcal diseases and for screening agents capable of antagonizing  
 or inhibiting expression of the protein -

Claim 1; Page 83; 108pp; English.

AAV81501 to AAV81679 represent specifically claimed protein sequences  
 isolated from Streptococcus pneumoniae. AAV81407 to AAV81590 represent  
 specifically claimed nucleotide sequences isolated from S. pneumoniae.  
 The sequences have antibacterial and antiinflammatory properties.  
 The protein sequences, and fragments of them, are useful as immunogens  
 and/or antigens. The nucleotide sequences can be used in vaccines and in  
 diagnostic assays. The proteins and nucleotides can be useful for the  
 detection and diagnosis of S. pneumoniae. The protein sequences are also  
 useful for screening an agent capable of antagonising, inhibiting or  
 interfering with the function or expression of the proteins in which the  
 agent is useful for treatment or prophylaxis of S. pneumoniae infection  
 and meningitis. AAV81591 to AAV81614 represent primers used in the



CC exemplification of the present invention.  
 XX  
 SQ Sequence 1237 AA;

alignment\_scores:  
 Quality: 329.00 Length: 532  
 Ratio: 1.061 Gaps: 21  
 Percent Similarity: 58.271 Percent Identity: 25.376

alignment block:

US-09-668-119-1 x AAY81609 ..

Align seq 1/1 to: AAY81609 from: 1 to: 1237

```

225 GCGCTCGTGGGGCCGGACAGTCTCGGGGAGTGGAGTTCGTCG 274
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
675 AlaserAlaserAlaserTherSerAlaserGluSerAlaserTherSerAl
275 CCATGGGACAGCCAAATGCTCTCTCAGGGGAGCGCTCTGGAGCTCG 324
:|||||:|||||:|||||:|||||:|||||:|||||:
691 aserAlaserAlaserTherSerAlaserAlaserAlaserTherSerAlas
325 GGGATGGCCCCCTACAGCATGGCTCTGCTGTACGGGCACTCCACAGAC 374
:|||||:|||||:|||||:|||||:|||||:|||||:
708 erAlaserAlaserThr.....SerAlaserGlu 717
375 CCAGCTCAGCTCCAGCAGGTGGGCTGTCAGCAGCAGCAGCAGCAGCAGC 424
|||||:|||||:|||||:|||||:|||||:|||||:
718 SerAlaserTherSerAlaserAlaserTherSerAlaserGlu 734
425 AGTTCCAGCAGCAGCAGCAGCGGCTACAGCAGCAGCAGCAGCAGCAGC 474
:|||||:|||||:|||||:|||||:|||||:|||||:
734 rAlaserTherSerAlaserAlaserTherSerAlaserAlaser 751
475 CAACAGCAGCAGGTCCAGGCTCAGCAGAGTCCAGCAGCAGCAGCAGTCCA 524
|||||:|||||:|||||:|||||:|||||:|||||:
751 laserTherSerAlaserGlySerAlaserTherTherSerAlaserAla 767
525 AGCAGTAGTGACAGCAGCAGCAGTCCAGCAGCAGCAGCAGCAGCAGCAG 574
|||||:|||||:|||||:|||||:|||||:|||||:
768 SerTherSerAlaserAlaserTherSerAlaserAlaserAlase 784
575 AGCATCTAATTAATTCATCATCAATAATCAGCAACAGATACAGCAGCAG 624
:|||||:|||||:|||||:|||||:|||||:|||||:
784 rAlaserAlaserGluSerAlaserTherTherAlaserGluSerAlaser 801
625 CAACAGCAGCAGCAGCAGCAGTCCAGCAGCAGCAGCAGCAGCAGCAGCAG 674
|||||:|||||:|||||:|||||:|||||:|||||:
801 hrSerTherSerAlaserTherTherSerAlaserGluSerAlaserThr 817
675 GCAGCAGCAGCAGCAGCAGCA...GCAGCAGCAGCAGCAGCAGTTCGAGGCC 721
|||||:|||||:|||||:|||||:|||||:|||||:
818 SerAlaserAlaserTherTherSerAlaserAlaserAlaserTher 834
722 AGCCACCAATTCAGCAGCAGCAGTTCAGCAGCAGCAGCAGCAGTTC 762
|||||:|||||:|||||:|||||:|||||:|||||:
834 rAlaserAlaserAlarg...GluAlargArgProGluProValAlargLeu 850
763 .....CCGCGCTCCAGGCTCTGCGCCAGCAGT 791
|||||:|||||:|||||:|||||:|||||:|||||:
851 AsnAlargInProValAlargInProGluInValAlargLeuAlargInLe 867
792 GCAG...CAGATGCATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 831
|||||:|||||:|||||:|||||:|||||:|||||:
867 uGlnInArgValAlargLeuGlnInArgProValAlargLeuG 884
832 .....CAGCCCAAGCAGCAGTCCAGTCCAGCAGCAGCAGCAGCAGCAG 864
|||||:|||||:|||||:|||||:|||||:|||||:
884 InArgInProValAlargInLeuGlnInValAlargLeuGlnIn 900
865 CCATGACAA...CTCCGCGCAGCAGTCCAGCAGCAGCAGCAGCAGTTCGTCG 911
|||||:|||||:|||||:|||||:|||||:|||||:

```

```

901 HisGlnInValLeuGlnProGlnHisArgGlnValAlargLeuGln 917
912 GCGGCAAGCTCTCCCTGGACAATGTTGTATACCCACACCACTGAAT 961
|||||:|||||:|||||:|||||:|||||:|||||:
917 nAlaHisGlnHisLeuAsnGlnArg..... 926
962 TTGTCGAGCTCCGATGGTGGTCGACAGCCCCAGTCGACGCCAGCTG 1011
:|||||:|||||:|||||:|||||:|||||:|||||:
927 .....GlnAlaProGlnLeuGlnInValAlargLeuGlnIn 941
1012 CAGCAG.....CAGCAGCAGCAGTACAGCAGCAGTCAAG.... 1044
|||||:|||||:|||||:|||||:|||||:|||||:
942 ArgGlnValAlargInProGlnInValAlargLeuHisGlnInLeu 958
1045 .....GTCGCCAGATGGTGGCTCCGAGTCCAGTCA 1078
|||||:|||||:|||||:|||||:|||||:|||||:
958 nArgValHisArgLeuArgArgInProValHisGlnSerGlnInVal 975
1079 GCCAGAGCAGCTCCCAATGCTCTGCTGCGCG.....TCACCG 1116
:|||||:|||||:|||||:|||||:|||||:|||||:
975 roValAlargInLeuProHisGlnInValAlargLeuGlnInAlaPro 991
1117 GGCACAGAGTGCAGACCCCGCAGTCGATGCCCTCCCGCCAGCAGCTG 1166
:|||||:|||||:|||||:|||||:|||||:|||||:
992 ValArgArgLeuGlnInValAlargLeuAlaProGlnProGlnPro 1008
1167 CCGGACGCGCGCGCAGCAGCTCAGCAGCAGCAGCAGCAGCAGCAGCTG 1216
|||||:|||||:|||||:|||||:|||||:|||||:
1008 lArgInProGlnInValAlargLeuGlnArgHisGlnInValAlarg 1025
1217 GCGCTCCGCGCAGTCCAGTCTCTGCTGCGCGCCCTCAGCAGCAGCAG 1266
|||||:|||||:|||||:|||||:|||||:|||||:
1025 rArgProLeuGlnInValLeu.....AlaProGlnPro 1035
1267 TCCGACAGCCAGTGCAGCGCGCGACCCACAGAACTTCAGTCCCTC 1316
|||||:|||||:|||||:|||||:|||||:|||||:
1036 GlnArgGlnInValHisArgLeuGlnArgGlnAlargLeuAla 1052
1317 ACCTGACCTTTAAACACACCTGTGAACCC...AGCTGTGCATGAGCC 1363
|||||:|||||:|||||:|||||:|||||:|||||:
1052 G.....HisGlnArgValArgProLeuGlnInValAlarg 1065
1364 CAGCTGGCTCCAGCAGCAGTGCAGCAGCAGCAGCAGTTCGACAGCTGA 1413
|||||:|||||:|||||:|||||:|||||:|||||:
1065 roGlnProGlnAlargGlnInValHisArgLeuGlnHisGlnInVal 1081
1414 CAGCTGTCAATGATC.....GAGCCCTGCGCGCAGTATGACACA 1457
|||||:|||||:|||||:|||||:|||||:|||||:
1082 ProLeuGlnInValAlargProGlnAlargGlnInValHisArg 1098
1458 GATCGACAAAGACAGCAAGAAAGAGCAGTGAAGTGAAG.... 1503
|||||:|||||:|||||:|||||:|||||:|||||:
1098 GlnArgGlnInValAlargLeuSerGlnHisGlnInValAlargGln 1115
1504 .....AGCTTCTGACAT.....CTGACAGACCC 1530
|||||:|||||:|||||:|||||:|||||:|||||:
1115 roGlnInAlaHisGlnLeuLeuAsnHisGlnInProValAlargInPro 1131
1531 TCGAAGCG.....TGTCCCTGACAGCCTTGA 1559
:|||||:|||||:|||||:|||||:|||||:|||||:
1132 GlnHisArgInAlaProGlnLeuGlnInValAlargInProValAlarg 1148
1560 AAGATGTGATC...GCCCTGGAGAA.....CTCAGAAATGATCAG 1600
|||||:|||||:|||||:|||||:|||||:|||||:
1148 nArgArgGlnInValAlargArgLeuGlnInValAlargInProGln 1165
1601 GGTGCCCATC.....CCGACGCGCGCGTGCAGCAGCAGCAGCAGCAG 1638
|||||:|||||:|||||:|||||:|||||:|||||:
1165 InValProValAlargInProGlnAlargGlnInValAlargArgPro 1179

```

seq\_name: /STD2/gcgdata/geneseq/geneeqp/AA2001.DAT: AAB65609

seq\_documentation\_block:



ID	AA869609 standard; Protein; 121 AA.
XX	
AC	AA869609;
XX	
DT	30-APR-2001 (first entry)
XX	
DE	Huntingtin accumulation inhibitor peptide HD-Q104-GFP.
XX	
KW	Neurological disorder; Huntington's disease; Alzheimer's disease;
KW	Parkinson's disease; prion disease; frontotemporal dementia;
KW	amyotrophic lateral sclerosis; spinal and bulbar muscular atrophy;
KW	dentatorubral-pallidoluysian atrophy; spinocerebellar ataxia type 1;
XX	SCA2; SCA3; SCA4; SCA5; SCA6; SCA7; protein accumulation; intrabody.
XX	
OS	Synthetic.
XX	
PN	WO200106989-A2.
XX	
PD	01-FEB-2001.
XX	
PF	24-JUL-2000; 2000WO-US20131.
XX	
PR	27-JUL-1999; 99US-0146047.
PR	21-JUL-2000; 2000US-0620955.
XX	
PA	(HUST/) HUSTON J S.
PA	(MESS/) MESSER A.
XX	(LECE/) LECERF J.
PI	Huston JS, Messer A, Lecerf J;
DR	WPI: 2001-182700/18.
XX	
PT	Inhibiting intracellular polypeptide accumulation, useful for treating
PT	neurological disorders, e.g. Alzheimer's disease, comprises contacting
PT	the polypeptide with a specific intrabody -
XX	
PS	Disclosure: Page 98; 108pp; English.
XX	
CC	The present invention describes a method for inhibiting the formation of
CC	aggregates of certain proteins, involving contacting the protein with a
CC	blinding molecule known as an intrabody. Proteins to be bound include
CC	those associated with neurological disorders, and so the method can be
CC	used in the prevention of diseases such as Alzheimer's, Parkinson's and
CC	Huntington's diseases, prion disease, frontotemporal dementia,
CC	amyotrophic lateral sclerosis, spinal and bulbar muscular atrophy,
CC	dentatorubral-pallidoluysian atrophy, spinocerebellar ataxia type 1
CC	(SCA1), SCA2, SCA3, SCA4, SCA5, SCA6 and SCA7.
XX	
XX	Sequence 121 AA;
XX	

```
alignment_scores:
  Quality: 328.50
  Ratio: 4.056
  Percent Similarity: 72.973
  Length: 111
  Gaps: 1
  Percent Identity: 69.369
```

```
alignment_block:
```

US-09-668-119-1 x AAB69609 . . .

Align seg 1/1 to: AAB69609 from: 1 to: 121

[illegible]

```

526 GCATCTAGTCGACGACGACGAGCTCCACGACGACGACGACGACGA 575
      |||||
68 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 84
576 GCATCTATTAAATTCATCATCAAAATCGACACAGATACGACGACG 625
      | |||||
84 n.....GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 94
626 AACGACGCTGCACGCCAATATGACGACGCTGCAGCTCCAAACAGCAACAG 675
      ||||| |||||
94 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 110
676 CACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 708
      ||||| |||||
111 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 121

```

```

676 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 708
      |||||
111 GInGInGInGInGInGInGInGInGInGIn 121

```











```

seq_name: /cogn2_6/prodata/2/1aa/5a_comb.pep:us-07-906-349A-6
seq_documentation_block:
; Sequence 6, Application US/07906349A
; Patent No. 5434064
;
; GENERAL INFORMATION:
; APPLICANT: Schlessinger, Joseph
; APPLICANT: Skolnik, Edward Y.
; APPLICANT: Margolis, Benjamin L.
; TITLE OF INVENTION: A NOVEL EXPRESSION-CLONING METHOD FOR
; TITLE OF INVENTION: IDENTIFYING TARGET PROTEINS FOR EUKARIOTIC TYROSINE KINASES
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/906,349A
; FILING DATE: 30-JUN-1992
; CLASSIFICATION: 435
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/643,237
; FILING DATE: 18-JAN-1991
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
;
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 801 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-07-906-349A-6

```

208 nTylValThrProValAsnArgAsnValCysAlaAlaAlaAlaGlyAla 225

1500 CATCTTACTCAAGTCCCT.....TTTTCCTGCTTCCTCTTGCTCATGC 1457

225 IaThrThrAlaAlaAlaCysCysCysAlaCysAlaAlaGlyCysThrGly 241

1456 TCTTGATCATGC.....GGCCACAGGCGCTGAGTGAATCCTC 1422

242 CysCysThrCysThrGlyAlaCysAlaGlyCysAlaGlyCysCysThrGly 258

1421 GACAGCTGCTTCAGCTGTGTCAGTACTGC..TGCCTCCAGCCTGCGTGG 1373

258 yhnhrGlyAlaGlyGlyGlyAlaGlyAlaGlyThrGlyCysAlaGlyAlaAlaCysA 275

1372 ACCCACTGGGCTCATGACACAGAGCTGGGGCTCCACAGGAGTAAAGCT 1323

275 IaCysCysGlyThrThrThrThrThrThrCysThrThrAlaAlaThrThrThrGly 291

1322 CCAGGTGAGGGAGACACTGAAATGTTCTGTGGGGTCCGCGCGCTCACTGGGCT 1273

292 GlyGlyThrGlyThrThr.....CysGlyAlaCysGlyGlyAlaGlyAla 306

1272 CTGGGAGGCG...TGGGTGAGGGCGCTGGCAGAGACTACTGGGAATG 1226

306 acYThrThrGlyThrCysGlyThr..... 313

1225 GGGGAGGCGCAAGCGTGAAGCTTGGAGTTGGGCTGACCTGGCGGCGC 1176

314 ..CysGlyGlyAlaCysAlaCysAlaCysThrCysCysCysThrCysAlaCysGly 329

1175 GCGTCGGGAGCGGCTGGGGGGAGGGGAGCATGCAGCTCGGGGCTGCAC 1126

330 ThrCysThrThrGlyThrGlyThrGlyThrGlyCysCysGlyGlyGly 346

1125 CTGCTGGCCCGTGAAGGGAGAGACAGCATGGGAGCGCTGCTGGCGTA 1076

346 rCys..... 347

1075 CCTGGACTCCGGAGCCACCATCTGGGACGCTGACCTGTCTGT..... 1032

348 .....AlaCysCysThr 352

1031 ...ACTGCTCTGCTGCTGC...TGCACCTGGGGTGCAC..... 996

353 GlyThrGlyAlaCysCysCysThrCysAlaCysThrThrThrGly 369

995 .....GGGGGCTGTCACCAACATCGGAGCTCGGACAAT 959

369 yGlyThrThrGlyGlyAlaAlaCysThrThrAlaGlyGlyGlyGly 386

958 TCACTGGTGGTGGGTATACACATTTGTCAGAGGAGACTCGGCGCTGT 909

386 LyrnhrGlyGlyAlaGlyGly.....GlyGlyGlyCysAlaCys 399

908 GACACCAAGGCTGGGCTCGACTCGAGCGGGAGGTGTGATGTTGTT 859

400 .....CysGlyGlyCysCys..... 404

858 CTGAGCAATCGAGAGCTGCTGGGGCTGTGTGCTGGCTGTGCTGTG 809

405 .....CysAlaGlyThrGlyGly..... 411

808 TGTGATGCTGTGCTGAGCGTGGGGGAGAGCTGGGAGGCGGAGGC 759

412 .....AlaCysAlaCys.....ThrGlyGlyGly 419

758 TGTGCG.....TGTGCTGATGGGTGCTGCTGAAT 730

420 AlaGlyAlaGlyThrGlyAlaAlaAlaCysCysAlaAlaCysCysThrTh 436

729 TGGTGGCTGGGCGTCCAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 680

436 rGly.....AlaAlaAlaThrCysCysCysCysGlyAlaCysCysCys 451



```

679 GCTGCTGTGCTGTGT..... 663
451 hrCysCysCysCysCysGlyGlyThrThrGlyAlaThrThrThra 467
662 .....TGAGCTGCAGCTGTGCTATTCGTCGACGCTGCTGTG 625
468 AlaAlaAlaAlaThrGlyCysAlaAlaAlaAlaCysThrThrAla 484
624 CTGC..... 621
484 scysThrAlaThrAlaAlaAlaThrThrAlaAlaGlyAlaAlaGly 501
620 .....TGCTGTATCTGTGCTGATTTGA 597
501 lyThrThrThrThrAlaThrThrAlaCysAlaAlaAla..... 514
596 TGATGCAATTATAGATGCTGCTGCTGC..... 564
515 .....ThrThrThrCysAlaAlaCysCysThrAlaAlaAlaThr 528
563 .....TGCTGC..... 558
528 rThrThrAlaCysGlyGlyThrThrThrThrGlyAlaAlaThrGly 545
557 .....TGCTGAGCTGCTGCTGC 540
545 lyAlaThrAlaThrThrAlaAlaAlaThrThrCysThrThrCys 561
539 TGCTGC.....ACTACTGCTTGGAACTG 517
562 AlaAlaAlaAlaAlaThrAlaAlaThrGlyThrThrThrAlaAla 578
516 CTGCTGCAGTGCCTGCTGAGCTGAGCTGAGCTGCTGCTGCTGCTG 474
578 rCysAlaCysThrGlyCysThrGlyCysThrCysThrCysThrThr 595
473 ..TGCTGCTGCTGCTGCTGCTGAGCGCCGCTGCTGCTGCTGCTG 435
595 hrCysCysCysCysThrCysCysThrThrThrGlyCysThrThrThr 611
434 .....TGCTGAGCTGCTGCTG.....TG 418
612 ThrThrThrThrThrCysAlaThrCysCysThrThrThrThrThr 628
417 TTGCTGCTGCTGCTGCTGAGCGCCAGCTGC.....TGAGCTGAGCT 377
628 strCysThrThrCysThrGlyThrCysAlaGlyThrGlyAlaCysGly 644
376 GGGCTGTGAGATTGCCGTAGACAGACAGCCATGCTGAGGGCCATC 327
645 ..AlaCysGlyAla.....GlyGlyAla 651
326 CCCGAGTCCAGAGAGCGGCTGCCCTGAGAGACATTCGCTGCCAT 277
652 GlyAlaAlaAlaGlyGlyGlyAlaGlyGlyAlaAlaAlaCysAla 668
276 GGCACCAAGCTACCCATCCGCCAGAGACTGTCGGGGCCCGAGAG 227
668 yAlaAlaAlaAlaAlaAla.....AlaAlaGlyThrAlaGlyG 681
226 GCATGCAATTCCAGCGGCTCCCGAGAGAGT 195
681 ly.....AlaAlaAlaAlaAlaAlaGly 688

```

seq\_name: /cgn2\_6/ptodata/2/iaa/bb\_comb.pep:us-08-916-352-2

seq\_documentation\_block:

; Sequence 2, Application us/08916352

; Patent No. 6166191

; GENERAL INFORMATION:

; APPLICANT: CHIRON CORPORATION

; TITLE OF INVENTION: HUMAN POLYHOMEOPTIC 1 (hph1) ACTS AS A

```

TITLE OF INVENTION: TUMOR SUPPRESSOR
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: CHIRON CORPORATION
STREET: 4560 HORTON STREET
CITY: EMERYVILLE
STATE: CA
COUNTRY: USA
ZIP: 94608
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/916,352
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: POTTER, JANE
REGISTRATION NUMBER: 33,332
REFERENCE/DOCKET NUMBER: 1355.
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-923-2707
TELEFAX: 510-655-3542
INFORMATION FOR SRO ID NO: 2
SEQUENCE CHARACTERISTICS:
LENGTH: 1004 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-916-352-2

```

#### alignment\_scores:

Quality: 389.00 Length: 522  
Ratio: 1.379 Gaps: 20  
Percent Similarity: 54.023 Percent Identity: 28.544

#### alignment\_block:

US-09-668-119-1 x US-08-916-352-2 ..

Align seg 1/1 to: US-08-916-352-2 from: 1 to: 1004

```

178 CAGAGCTGACTGGCGGAGCTGCTGCGGAGCCGCTGGAATTGGCAGTCC 227
|||||.....:|||||.....:|||||.....:|||||.....:
260 GlnSerLeuAsnLeuSerGlnAlaGlyGlyGlySerGlyAsnSerIlePr 276
228 T.....CCTCGGGGCGCGGAGACAGTCTCTGGGGGAGTGGTGA 265
|.....:|||||.....:|||||.....:|||||.....:
276 OGlySerMetClyProGlyGlyGlyGlyGlnAlaAlaGlyGlyGly 293
266 GCTTGTGTCATGGAGACACCAATGCTCTCTCAGGGCAGCCGCTCT 315
:.....:|||||.....:|||||.....:|||||.....:
293 InLeuProSerSerGly.....MetGlyGlySerCysProAlaGlys 307
316 GGCAGCTCGGGGAGTGGCCCTTCACAGCATGGCTGTGCTACGGCAC 365
|||||.....:|||||.....:|||||.....:|||||.....:
308 GlyThrGlyAlaValGlnProLeuProAlaAlaGlnThrValThrVal 324
366 TCACAGACCCAGCTGAGCTCCAGCTCCAGAGTGGGCTGACAGCAGCAG 415
:.....:|||||.....:|||||.....:|||||.....:
324 rGlnGlySerClnThrGlnAlaGlnSerAlaAlaAlaAlaGlyAlaGln 341
416 AACAGCAGCAGTTCACAGCAGCAGCAGCAGCGGCGCTACAGCAGCAG 465
:.....:|||||.....:|||||.....:|||||.....:
341 LaaspGlySerGlyGlnGlnAsnValGlyMetAsnLeuThrArgThrAla 357
466 CACAGCAGCAGCAGCAGCTCCAGGCTCAGCAGCAGTGCATGCAGCA 515
:.....:|||||.....:|||||.....:|||||.....:
358 ThrProAlaProSerGlnThrLeuIleSerSerAlaThrTyThrGlnI 374

```











1596 CATGGGTGC.....CCACTCCCCACGGCCCCGGTGCAC 1633  
 846 .....CysProSerLeuGlyProValProHisArgProPro...PheG 859  
 1634 GACCAACACAGCTACTATGACGCGCGCTCGATGCGCTCTGGCCA 1683  
 859 LuProGlySerAlaValAlaThrValProProTyrLeuGlyProAsp 875  
 1684 ACATCCCTGACCTGCTTCACACCATTCCTGTACCGACATTCGTTCC 1732  
 875 rProAlaLeuArgThrLeuSerGlyTyrAlaArgProHisValMetSer 891  
 seq\_name: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:US-08-991-300-2

seq\_documentation\_block:  
 ; Sequence 2, Application US/08991300  
 ; Patent No. 5973225

GENERAL INFORMATION:

APPLICANT: D'OVIDIO, RENATO

APPLICANT: PORCEDDU, ENRICO

APPLICANT: MERCHITELLI, CINZIA

APPLICANT: CARDELLI, LUISA ERCOLI

TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF A GENE

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

STREET: 1755 S. JEFFERSON DAVIS HIGHWAY

CITY: ARLINGTON

STATE: VA

COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/991,300

FILING DATE: 16-DEC-1997

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: IT MI 96/A 002663

FILING DATE: 19-DEC-1996

ATTORNEY/AGENT INFORMATION:

NAME: OBLON, NORMAN F.

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 2264-0201-0X

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-413-3000

TELEFAX: 703-413-2220

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 369 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-991-300-2

alignment\_scores:  
 Quality: 311.50 Length: 377  
 Ratio: 1.639 Gaps: 16  
 Percent Similarity: 50.398 Percent Identity: 32.095

Alignment block:  
 US-09-668-119-1 x US-08-991-300-2 ..

Align seg 1/1 to: US-08-991-300-2 from: 1 to: 369

400 CTGCAGCAGCAGCAGCAGCAGCAGCTTCCAGCAGCAGCAGCAGCAGC 449  
 6 LeuGluArgProSerGlnGlnGlnProLeuProProGlnGlnThrLeu 22  
 450 GCTACAGCAGCAGCAGCAGCAG.....CAGCAGCAACAGCAGCAGCTCCAGG 493  
 22 rHisHisHisGlnGlnGlnProIleGlnGlnGlnProHisGlnPhePro 39  
 494 CTGCAGCAGCTCCATGCAGCAGCAGCTTCCAGCAGCTAGTGCAGCAGG 543  
 39 LngLngInProCysSerGlnGlnGlnGlnGlnProProLeuSerGlnGln 55  
 544 CAGCAG.....CTCCAGCAGCAGCAG.....CAGCAGCA 572  
 56 GlnGlnProProPheSerGlnGlnGlnGlnProProPheSerGlnGln 72  
 573 GCAGCATCTAATTAAATTGCATCATCA...AATCAGCAAGATA.... 615  
 72 ngInProValLeuProGlnGlnProSerPheSerGlnGlnLeuPro 89  
 616 .....CAGCAGCAGCAGCAGCAGCTCAGCAGTATGACAGCTGCAG 657  
 89 rPheSerGlnGlnGlnGlnProProPheSerGlnGlnGlnProVal 105  
 658 CTCCAGCAGCAGCAGCAGCAGCAGCAGCAG.....CAGCA 698  
 106 LeuProGlnGlnProSerPheSerGlnGlnGlnProProPheSerGln 122  
 699 GCAGCAGCAGCAGCTTGGAGCCCGCCAGCCCAATT.....CAGCAGCC 742  
 122 ngInLeuProProPheSerGlnGlnGlnProValLeuProGlnGlnPro 139  
 743 CGATGCAGCAGCAGCAGCAGCTCCGCCC...TCCAGCAGCTGCCCGCAG 789  
 139 rPheSerGlnGlnGlnGlnProProPheSerGlnGlnLeuProPhe 155  
 790 CTGCAGCAGCTGCATCAGCAGCAGCAGCCCGCCAGCAGCCCA 839  
 156 SerGlnGln.....GlnGlnProValLeuProGln 165  
 840 GCAGCCTCCAGTTGCTCAGAACCAACCACTCCGCCAGCAGCTGC 889  
 165 ngInProProPheSerGlnGlnGlnGlnProIleProGln..... 180  
 890 AGACCCAGCCTTTGGTGCACAGCGCAAGCTTCCCTGGACAATGTTG 939  
 181 .....GlnProProPheSerGlnGln.....GlnProVal 191  
 940 TATACCAACCAACCACTGAATTGTCAGCTCCGATGAGTGAGCA 989  
 192 LeuLeuGlnGlnGlnGlnProPheValHisProSerIleLeuGlnGln 208  
 990 GCCCCAGTGCAGCCCGAGTGCAGCAGCAGCAGCAGCTGACAGCAG 1039  
 208 uAsnProCysLysValPheLeuGlnGlnGlnGlnCysSerProTyrAlaMet 225  
 1040 CTCAGGCTGCCAGAGGTGGCTCCCGAGTCCAGGTCCAGCCAGCAGCAG 1089  
 225 rGlnSer.....LeuAlaArgSerGlnMetLeuGlnGlnSer 238  
 1090 CTCCCATGCTGTCCTCGCGGTCAACCGGCGCAGCAGGTGAGACCCGCA 1139  
 239 CysHisValMetGlnGlnGlnGlnGlnGlnGlnLeuProGlnIleProGln 255  
 1140 G..... 1140  
 255 ngInSerArgTyrGlnAlaIleArgAlaIleValTyrSerIleIleLeuG 272  
 1141 .....TCGATGCCCTCCCTCCCGCCAGCGCTCC 1167  
 272 LngLngLngLngLngLngValGlnGlnGlnGlnGlnGlnGlnPro 288  
 1168 CCGCAGCCCGGCCAGCCAGCTCAGCAGCC.....AATCCAGCAGCTCAG 1211











```
429 CCAGCAG.....CAGCAGAGCGCGCTACAGCAGCAGCAGCAGC 469
|||||
452 oGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 469
|||||
470 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 519
|||||
469 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 485
|||||
520 TTCCAGCAGTAGTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 569
|||||
486 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 502
|||||
570 GCAGCAGCATCTAATTAAATTCATCAATAATCAGCAGCAGCAGCAG 619
|||||
502 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 511
|||||
620 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 669
|||||
511 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 520
|||||
670 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 719
|||||
521 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 537
|||||
720 CCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 766
|||||
537 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 554
|||||
767 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 813
|||||
554 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 570
|||||
814 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 860
|||||
571 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 587
|||||
861 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 910
|||||
587 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 604
|||||
911 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 960
|||||
604 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 611
|||||
961 TTGTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1010
|||||
612 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 623
|||||
1011 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1060
|||||
623 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 637
|||||
1061 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1110
|||||
638 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 650
|||||
1111 TCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1160
|||||
651 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 667
|||||
1161 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1204
|||||
667 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 684
|||||
1205 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1254
|||||
684 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 700
|||||
1255 TCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1304
|||||
701 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 716
|||||
```

```
1305 CAGTCCTCCCTCAGCTGACCTTTAAACACACCTGTGAACCCAGCTCTG 1354
716 .....
1355 TCATGAGCCAGCTGCTCAGCCAGCAGCAGCAGCAGCAGCAGCAGCAG 1404
717 .....
1405 AGCCTAAGCAGCTGTCAGATCAGATCAGCAGCAGCAGCAGCAGCAGCAG 1454
728 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 743
1455 CAGATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1504
744 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 760
1505 GCCTTGGACATTCGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1554
760 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 773
1555 TTGCAAAAGTGTGAGATCGCCCTGAGAGAACTCAAGAA 1592
773 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 785

seq_name: /cgn2_6/plodata/2/iaa/5A_COMB.pep:US-08-209-747-2
seq_documentation_block:
; Sequence 2, Application US/08209747
; Patent No. 5733771
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Coldin, Mark
; TITLE OF INVENTION: CDNA Encoding Minor Ampullate Spider
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-3487
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/209,747
; FILING DATE: 14-MAR-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-104P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 832 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: Internal
; ORIGINAL SOURCE:
; ORGANISM: N. clavipes
; TISSUE TYPE: minor ampullate gland
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..309
```



US-08-209-747-2

alignment\_scores:

Quality:	293.00	Length:	499
Ratio:	1.118	Gaps:	17
Percent Similarity:	52.505	Percent Identity:	29.259

alignment\_block:  
US-09-668-119-1/rev x US-08-209-747-2

Align seq 1/1 to: US-08-209-747-2 from: 1 to: 832

```

1719 GGTACAGGAGGATGTTGAAGACAGTGAAGGATGTTGGCAGGAGCGCA 1670
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
133 GLTYRGLYRGLYRGLYRGLYRGLYRGLYRGLYRGLYRGLYRGLYRGLY 149
1669 TCCAGAGCGGCTGCGATGACTGCTGTTGGTCGTCGCGACGGGGG 1620
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
149 YALAGLYALAGLYALAGLYALAGLYALAGLYALAGLYALAGLYAL 165
1619 CGGTGGGGGAGTGGGCGACGATGCTGATGCTTCTCCAGGCGCAT 1570
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
165 AGLYALAGLYALAGLYALAGLYALAGLYALAGLYALAGLYALAGLY 182
1569 CTCACACTTTTGCAGGCTTTCAGGGGACACGCTTCGAGGGGCTGTCA 1520
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
182 LAGLYGLTYRGLYRGLYRGLYRGLYRGLYRGLYRGLYRGLYRGLY 195
1519 GAATGTCACAGGCTCTCATCTCTACGCTGCTTTTCTGCTCTCG 1470
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
196 ALAALAGLYALAGLYALAGLYALAGLYALAGLYALAGLYALAGLY 206
1469 TTCTTTCGATCTGTTGATCATGCGCGGCGGCTCGATGTACTTCA 1420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
207 .....GlnGlyGLYRGLYRGLYRGLYRGLYRGLYRGLYRGLYRGLY 221
1419 CAGCTGCTTCACTTTCAGGATGACTGCTCTCTCAAGCTGGCTGGAGC 1370
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
221 LAALAGLYALAGLYALAGLYALAGLYALAGLYALAGLYALAGLYAL 231
1369 CAGCTGGGCTCATGACAGAGCTGGGCTTCAAGGTGTTAAAGTCCA 1320
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
232 .....GlyTYRGLY..... 234
1319 GGTGAGGGGACATGAAGTTCTGTGGGCTCCGCCGCTCACTGGGCTTG 1270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
235 ....ArgGLYALAGLYALAGLYALAGLYALAGLYALAGLYALAGLY 250
1269 GAGGGGCTGCGGTGAGGGCTGGGAGGAGCTGAGGATGAGGGCAG 1220
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
250 LY.....AlaGLYSerTYRGLYRGLYRGLYRGLYRGLYRGLYRGLY 259
1219 GGCACAGAGTACGTTGGATTGGGCTGTGAGCTGGGCTGGCGGCTGC 1170
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
260 GLY.....AlaGLYALAGLYAL 265
1169 GGGGACGGCT..... 1160
265 aclylaalaalaalaalaalaalaalaalaalaalaalaalaalaala 282
1159 .....GGGGGGGAGGGG 1148
282 *****GlyAlaGLYALAGLYALAGLYALAGLYRGLYRGLY 298
1147 GCATGCACTGCGGGGCTGACACTGCTGCCCGGTGACGCGAGGACAC 1098
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
299 GLTYRGLYALAGLY.....AlaGLYALAGLYALAGLYALAGLYALAGLY 314
1097 ATGGGAGGCTGCTGCTGCTGACCTGACCTCGGAGGACACACTGGGC 1048
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
314 aclylaaglylaaglylaaglylaaglylaaglylaaglylaaglyla 323

```

```

1047 AGCCTGAGCTGCTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 998
324 .....ArgGLYALAGLYALAGLYALAGLYALAGLYALAGLYALAGLY 333
997 CTGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 948
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
334 ALAGLYALAGLYALAGLYALAGLYALAGLYALAGLYALAGLYALAGLY 344
947 TGGGTATACACATTTGCTCCAGGAGAGCTTGGCGCTGACACCAAGG 898
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
345 GLYGLNserGLYTYRGLYALAGLY..... 353
897 CTGGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 857
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
353 LAGLYALAGLYALAGLYALAGLYALAGLYALAGLYALAGLYALAGLY 369
856 GAGCAACTGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 807
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
370 GLYARGLYALAGLYALAGLYALAGLYALAGLYALAGLYALAGLYALAG 386
806 TGATGCTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 767
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
386 YALAGLYALAGLYALAGLYALAGLYALAGLYALAGLYALAGLYALAG 731
766 .....GCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 731
403 LYALAGLYALAGLYALAGLYALAGLYALAGLYALAGLYALAGLYALAG 419
730 TTGGTGGCTGGGCTCCCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 681
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
420 GLYALAGLYALAGLYALAGLYALAGLYALAGLYALAGLYALAGLYAL 436
680 TGCAGCTGCTGCT.....GTTGTCGAGCTGCGAC 652
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
436 AGLYALAGLYALAGLYALAGLYALAGLYALAGLYALAGLYALAGLYAL 453
651 CTGCTGCTATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 602
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
453 LAGLYALAGLYALAGLYALAGLYALAGLYALAGLYALAGLYALAGLY 468
601 TTTGATGATGCAATTAAATGATGCTGCTGCTGCTGCTGCTGCTGCTG 552
469 .....GlyGLYTYRGLYRGLYRGLYRGLYRGLYRGLYRGLYRGLY 478
551 AGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 502
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
478 YALAGLYALAGLYALAGLYALAGLYALAGLYALAGLYALAGLYALAGLY 495
501 CTGCTGAGCTGGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 452
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
495 LYGLYGLNGLYGLYTYRGLYALAGLYALAGLYALAGLYALAGLYALAG 511
451 GCGCGGCTGCTGCTGCTGCT.....GGAAGCTGCTGCTGCT 417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
512 ALAGLYALAGLYALAGLYALAGLYALAGLYALAGLYALAGLYALAGLY 528
416 TGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 371
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
528 YALAGLYALAGLYALAGLYALAGLYALAGLYALAGLYALAGLYALAGLY 543

```

seq\_name: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:US-08-458-298-2

seq\_documentation\_block:

Sequence 2, Application US/08458298

Patent No. 5756677

GENERAL INFORMATION:

APPLICANT: Lewis, Randolph V.

APPLICANT: Coiglin, Mark

TITLE OF INVENTION: CDNs Encoding Minor Ampullate Spider

TITLE OF INVENTION: Silk Proteins

NUMBER OF SEQUENCES: 56

CORRESPONDENCE ADDRESS:

ADDRESSEE: Birch, Stewart, Kolasch & Birch















```

606 CTGATTTGATGATGATTAATTAATGATGCTGCTGCTGCTGCT 557
127 S.....CysSerLeuPheTrpCysCysCysCysCysLeuA 140
556 GCTGAGAGCTGC.....TGCTGCTGCTGACACTGCTTGGAGAC 519
140 snTrpHrLeuTrpTrpTrpGlyCysCysCysCysCys***LeuTrp... 155
518 TGCTGCTGATGACACTGCTGAGCGCTGAGACTGCTGCTGCTGCTG 469
156 .....TyrAsnCysCysCysCysCysCysCys 163
468 CTGCTGCTGCTGCTGCTGAGCGCGCTGCTGCTGCTGCTGAACTGC... 423
163 eCysCysCysCysCysCys.....CysCysCysHsTrpArgCysSers 177
422 .....TGCTGTTCCTGCTGC..... 408
177 eSerGlyLeuSerValPheHsCysCysCysArgCysTrp***ArgIle 193
407 ...TGCTGAGCGCGCACTGCTGAGCTGAGCTGAGCTG... 375
194 AspThrValAsnPheAsnCysTrpAspTrpAsnPheLeuHsGlnIle 210
374 .GTCGT.....GGAGTTCGCTGACACGACA 348
210 uValCysSerHsAsnIleValIleValIleThrThr 221

```

seq\_name: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:US-08-415-751-6

seq\_documentation\_block:

Sequence 6, Application US/08415751

Patent No. 5643772

GENERAL INFORMATION:

APPLICANT: PETERSEN, CAROLYN

APPLICANT: LEBCH, JAMES

APPLICANT: NELSON, RICHARD, C.

APPLICANT: GGT, JIRI

TITLE OF INVENTION: POLYPEPTIDES BINDING ANTI-

TITLE OF INVENTION: CRYPTOSPORIDIUM ANTIBODIES, DNA

TITLE OF INVENTION: AND RNA ENCODING THEM, HYBRID

TITLE OF INVENTION: VECTOR AND TRANSFORMED HOST AND

TITLE OF INVENTION: METHODS FOR IMMUNOTHERAPY AND

NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:

ADDRESSEE: PHILLIPS, MOORE, LEMPIO & FINLEY

STREET: 385 Sherman Avenue, Suite 6

CITY: Palo Alto

STATE: California

COUNTRY: United States of America

ZIP: 94306-1840

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Kb storage

OPERATING SYSTEM: DOS

SOFTWARE: Wordperfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/415,751

FILING DATE: 03-Apr-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/071,880

FILING DATE: June 1, 1993

APPLICATION NUMBER: 07/891,301

FILING DATE: May 29, 1992

ATTORNEY/AGENT INFORMATION:

NAME: Hana Dolezalova

REGISTRATION NUMBER: 30,518

REFERENCE/DOCKET NUMBER: 480.19-2 (HHD)

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 324-1677

TELEFAX: (415) 324-1678

```

; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 362 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; MOLECULE TYPE: linear
; ORIGINAL SOURCE:
; ORGANISM: Cryptosporidium parvum
; FEATURE:
; NAME/KEY: Positions coded by nonsense codons are
; NAME/KEY: identified as Xaa.
; US-08-415-751-6

```

```

alignment_scores:
  Quality: 288.00      Length: 189
  Ratio: 3.789        Gaps: 11
  Percent Similarity: 40.212  Percent Identity: 31.217

```

alignment\_block:

US-09-668-119-1/rev x US-08-415-751-6 ..

Align seg 1/1 to: US-08-415-751-6 from: 1 to: 362

```

891 CTGCACACTGTGGCGGAGATTGATGTTGTTCTGACGACACTGGAGCT 842
128 LeuArgThrTrp.....AsnTrpPheI 135
841 GCTGGGCTGTGTGTGGCGGCTGTGTGCTGTGTGTGTGTGTGTGTGTG 792
135 eIleArgVal*****ThrThrValIleValIleAsnHsSerCysV 152
791 ACCTGCTGGGCGAGAGCTGGAGGCGAGGCTGTGTGTGTGTGTGTGTG 742
152 aLeuIle.....Gln 155
741 TGCTGCTGATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGCT 693
156 HsLeuPheAsnTrp AspTrp.....CysA 164
692 .....TGCTGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 672
164 snPheAsnValThrAspIleArg***TrpCysArgCysCysHsPhe 180
671 TGCTGTTGTGGAGCTGACAGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 622
181 ValSerLeuGlySerSerSerCysArgTrpHsCysCysCysCysCysC 197
621 CTGCTGATCTGTGTGTGATTTGTGATGATGACATTAATTAATGATGCT 572
197 sCysCysPhe.....CysC 202
571 GCTGCTGCTGCTGCTGCTG...AGCTGCTGCTGCTGCTGCTGCTGCTGCT 525
202 yCysCysCysAsnTrpLeu***CysCysCysCysGlyTrp 218
524 TGAAGCTGCTGCTGATGACACTGTGTGTGTGTGTGTGTGTGTGTGTGT 475
219 TrpLeu.....ThrTrpAsnLeuCysCys 227
474 CTGCTGCTGC.....TGCTGCTGCTGACGCGGCTGCTGCTGCTGCT 434
227 sCysCysCysTrpPheLeuSerCysCysAsnAsp..... 239
433 GCTGAGACTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 384
240 ..TrpIleYsserCysGlyCysCysCysLeuArg***Trp*** 254
383 TGCAGCTGGGTGT 369
255 *****LeuValCys 259

```



```

seq_name: /cgn2_6/prodata/2/1aa/5A_COMB.pep:US-08-089-862-11
seq_documentation_block:
; Sequence 11, Application US/08089862
; Patent No. 5723588
;
; GENERAL INFORMATION:
; APPLICANT: Donofrio, David A
; APPLICANT: Stedronsky, Erwin R
; TITLE OF INVENTION: PROTEIN-ENRICHED THERMOPLASTICS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSER: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/089,862
; FILING DATE: 09-JUL-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A-58854/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
;
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1018 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-089-862-11

alignment_scores:
      Quality: 287.00      Length: 532
      Ratio: 0.966      Gaps: 18
      Percent Similarity: 55.827      Percent Identity: 27.820

alignment_block:
US-09-668-119-1/rev x US-08-089-862-11 ..
Align seg 1/1 to: US-08-089-862-11 from: 1 to: 1018

1719 GGTACAGGATGTTGAAGACAGGTGAGCGATGTGGCCAGGACGCA 1670
|||||
102 GYAlaGlySerGlyAlaGlySerGlyAlaGlySerGlyAlaGlySerGly 118
1669 TCCAGGAGG.....GCTGGCATAGTCTCTCTGTTGGTGGG 1629
|||||
118 YAlaGlyAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGlyAlaGly 134
1628 CACCGGGGCGGTGGGAGTGGGACGCGCATGTCATCTTGAGTTTC 1579
|||||
134 YSerGlyAlaGlyAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGly 151
1578 CAGGCGCATCTCACACTTTTGCAGGCTTCAGGGACCGCTTCGAGG 1529
|||||
151 TGGAlaSerIleYsValAlaValaSerAlaGlyProSerAlaGlyTyr 167
1528 GGTGTC.....TCAGATGCCAGAAAGGCTTCTATCTTACTCAGG 1488
|||||
168 GYAlaGlyAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGlyAlaGly 184

```

```

1487 TCCTTTTCTGCTCTGTTCTGTCGATCTTGTGATCATCGGCGCAG 1438
|
184 Y.....SerGlyAlaG 188
1437 GGGCTCGATGTAATTCGACAGCTGCTCA...GCTTTCACAGTACTCT 1391
|||||
188 YAlaGlySerGlyAlaGlySerGlyAlaGlySerGlyAlaGlySerGly 204
1390 GCTCTCAGCTGCTGGCGACGACGCTGCTCAGACGAGTGGGCTTC 1341
|||||
205 AlAGlyAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGlyAlaGly 221
1340 ACAGGTGCTTAAAGCTCCAGGTGAGGACACTGAAGTCTGTGGG 1291
|
221 TgYAlaAlaProGlyAlaSerIleYs.....ValAlaValaSerA 235
1290 CCGCGCCGCACTGGCTCTGTCGAGGCGTGGCGGTGAGGCG...TG 1244
|||||
235 lAGlyProSerAlaGlyTyrGlyAlaGlyAlaGlySerGlyAlaGly 251
1243 GGAAGCTACTGGAGATGGGCGAGGCCAGAGCTGACGTTGAGTTGG 1194
|||||
252 GlySerGlyAlaGlyAlaGlySerGly..... 260
1193 TGTGAGCTGGGCTGGCGGCTGGCGGAGACGCTGCGGGGAGGCGCAT 1144
|||
261 ....AlaGlyAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGly 276
1143 CGACTCGGGGCTGTCACACTGCTGCGCTGACCTGAGGAGGACGATG 1094
|||||
276 YSerGlyAlaGlyAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGly 292
1093 GGAGCTGCTGCTGCTGACCTGAGCTCGGAGCGCACCATCTGGGCGC 1044
|||||
293 AlAGlySerGlyAlaAlaProGly..... 300
1043 TGAGCTGCTGTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 994
|||||
301 ....AlaSerIleYsValAlaValaSerAlaGlyProSerAlaGly 316
993 GGGCTGCTGACACCAATCGAGCTCGGACAAATTTCAGTGGTGGG 944
|||||
316 YAlaGlyAla..... 319
943 TATACACATTTTCAGGAGAGCTGCGCTGCTGCTGCTGCTGCTGCTG 894
|||||
320 .....GlySerGlyAlaGlyAlaGlySerGlyAlaGly 330
893 GTCGCGACTGTGGCGGAGTT.....GTGATGTTGTTCTGAGC 853
|||||
330 YAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGlyAlaGlySerGly 347
852 AACTGAGGCGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 803
|||||
347 lAGlyAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGlyAlaGly 363
802 GCATCTGCTCAGCTGCTGGGCGAGCGTGGAGGCGGAGGCTGTGGC 753
|||||
364 GYAlaGlyAlaGlySerGlyAlaAlaProGly..... 374
752 TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 703
|||||
375 ....AlaSerIleYsValAlaValaSerAlaGlyProSerAlaGly 390
702 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 653
|||||
390 YAlaGlyAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGlyAlaGly 406
652 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 603
|||||
407 SerGlyAlaGlyAlaGlySerGlyAlaGlyAlaGly..... 418

```







```
301 ...AlaserIlelysvAlaIaValserAlaGlyProserAlaGlyTyrG 316
993 GGGCTGCTGCACACCATCGAGCTCGACAAATTTCAGTGTGTTGGC 944
316 lYAlaGlyAla..... 319
943 TATACACATTTCACAGAGAGCTTCGCTGTGACACCAAGCTGG 894
320 .....GlySerGlyAlaGlyAlaGlySerGlyAlaGly 330
893 GTCGCGAGCTGCGCGGAGTT.....GTGATGCTGCTTCGAGC 853
330 yAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGlyAlaGlySerGly 347
852 AACTGAGGCTGCTGGGCTGTGGGCTGTGGGCTGTGGTGTGAT 803
347 laGlyAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGlyAlaGlySer 363
802 GCATCTGCTGAGCTGCTGGGCGAGAGCTGGAGGCGGAGGCTGTGGC 753
364 GlyAlaGlyAlaGlySerGlyAlaAlaProGly..... 374
752 TGCTGCATCGGTGCTGCTGATGTGGCTGGCTCCAAAGCTGTCTG 703
375 ...AlaserIlelysvAlaIaValserAlaGlyProserAlaGlyTyrG 390
702 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 653
390 lYAlaGlyAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGlyAlaGly 406
652 GCTGTGCTATTGCTGCACTGCTGTGCTGCTGCTGCTGCTGCTGCTGA 603
407 SerGlyAlaGlyAlaGlySerGlyAlaGlyAlaGly..... 418
602 TTTGATGATGCAATTAAATTAGATGCTGCTGCTGCTGCTGCTGCTG 553
419 .....SerGlyAlaGlyAlaGlySerGlyAlaG 428
552 GAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGAC 503
428 lYAlaGlySerGlyAlaGlyAlaGlySerGly..AlaGlyAlaGlySer 443
502 TCTGCTGAGCTGGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGT 453
444 GlYAlaIaIaProGlyAlaSerIlelysvAlaIaValserAlaGlyProse 460
452 AGCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 403
460 rAlaGlyTyrGlyAlaGlyAlaGlySerGlyAlaGlyAlaGlySerGlyA 477
402 CAGCGCCACTGCTGAGCTGAGCTGGGTCTGTGGAGTTGCCGTAGACA 353
477 laGlyAlaGlySerGlyAlaGlyAlaGlySer..GlyAlaGlyAlaGlys 493
352 CGACAGCCATGCTGTGAGGGGCGCATCCCGAGTCCCGAGAGCGGCTGC 303
493 erGlyAlaGlyAlaGlySerGlyAlaGlyAlaGlySerGlyAlaGlyAla 509
302 CCTGAG...AGAGACATTGGCTGTCCATGGCACCAAGCTACCCATCCC 256
510 GlySerGlyAlaGlyAlaGlySerGlyAlaAlaProGlyAlaSerIlely 526
255 GCCCAGAGACTGTCCGGGCGCCGAGAGAGCATGCCAATTCACGCGCTC 206
526 svAlaIaValserAlaGlyProserAlaGlyTyrGlyAlaGlyAlaGlys 543
205 CCGCAGCAGGTCCGCGCATGAGCTCTGAGATGCAATTC 167
543 erGlyAlaGly..AlaGlySerGlyAlaGlyAlaGlySer 555
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 28, 2002, 17:49:47 ; Search time 86.87 Seconds  
(Without alignments)  
4536.338 Million cell updates/sec

Title: US-09-668-119-1

Perfect score: 1740

Sequence: 1 atgaggaagaagcgtgtgtgc.....cacatcgttcacgcatga 1740

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*

1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCUTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	181.2	10.4	3489	2	US-08-728-323A-1 Sequence 1, Appli
2	181.2	10.4	32207	2	US-08-770-379-20 Sequence 20, Appli
3	181.2	10.4	32207	4	US-08-757-669A-20 Sequence 20, Appli
4	171.8	9.9	543	6	5273901-6 Patent No. 5273901
5	170	9.8	533	6	5482709-5 Patent No. 5482709
6	142.6	8.2	397	3	US-09-253-691-3 Sequence 3, Appli
7	130.8	7.5	234	1	US-08-469-802B-3 Sequence 3, Appli
8	130.8	7.5	234	2	US-08-267-803B-3 Sequence 3, Appli
9	130	7.3	688	4	US-08-998-416-915 Sequence 915, App
10	126.6	7.3	6530	2	US-08-146-930-1 Sequence 1, Appli
11	126.6	7.3	6530	4	US-08-458-240-1 Sequence 1, Appli
12	126.6	7.3	6530	5	PCIT-US93-03993-1 Sequence 1, Appli
13	120.6	6.9	477	4	US-09-135-994-1 Sequence 1, Appli
14	118.8	6.8	2580	3	US-09-050-863-2 Sequence 2, Appli
15	118.8	6.8	5452	2	US-09-130-114-1 Sequence 1, Appli
16	118.8	6.8	9600	4	US-08-910-647-1 Sequence 1, Appli
17	118.8	6.8	10596	1	US-07-884-811-15 Sequence 15, Appli
18	118.8	6.8	10596	1	US-07-885-971-15 Sequence 15, Appli
19	118.8	6.8	10596	1	US-08-087-783A-15 Sequence 15, Appli
20	118.8	6.8	10596	1	US-08-194-088B-15 Sequence 15, Appli
21	118.8	6.8	10596	2	US-08-194-087-15 Sequence 15, Appli
22	118.8	6.8	10596	5	PCIT-US93-04648-15 Sequence 15, Appli
23	114.8	6.6	2793	1	US-08-209-747-1 Sequence 1, Appli
24	114.8	6.6	2793	1	US-08-458-298-1 Sequence 1, Appli
25	113.8	6.5	2214	4	US-09-043-303-7 Sequence 7, Appli
26	113.4	6.5	2214	4	US-08-864-038A-1 Sequence 1, Appli
27	113.4	6.5	3331	3	US-08-864-038A-2 Sequence 2, Appli

C	28	113.4	6.5	3331	3	US-08-864-038A-4	Sequence 4, Appli
	29	111.2	6.4	195	1	US-08-469-802B-2	Sequence 2, Appli
	30	111.2	6.4	195	2	US-08-267-803B-2	Sequence 2, Appli
C	31	110.8	6.4	786	1	US-08-403-379A-2	Sequence 2, Appli
	32	110.8	6.4	786	2	US-08-929-414-2	Sequence 2, Appli
	33	110.8	6.4	9551	1	US-08-056-200-93	Sequence 93, Appli
	34	110.8	6.4	9551	2	US-08-800-644-93	Sequence 19, Appli
	35	109	6.3	2190	4	US-09-625-188-19	Sequence 17, Appli
	36	105.8	6.1	165	4	US-09-103-429A-1	Sequence 1, Appli
	37	101.8	5.9	2455	4	US-08-781-891-208	Sequence 208, App
C	38	101.8	5.9	16442	3	US-08-425-069-3	Sequence 3, Appli
	39	98.8	5.7	1995	1	US-08-317-844B-3	Sequence 3, Appli
	40	98.8	5.7	1995	2	US-08-317-844B-3	Sequence 2, Appli
	41	97.4	5.6	2821	4	US-09-103-429A-2	Sequence 6, Appli
	42	96	5.5	154	1	US-08-469-802B-6	Sequence 6, Appli
	43	96	5.5	154	2	US-08-267-803B-6	Sequence 3, Appli
	44	93.6	5.4	336	2	US-07-814-220-3	Sequence 4, Appli
C	45	93.6	5.4	336	2	US-07-814-220-4	Sequence 4, Appli

#### ALIGNMENTS

RESULT 1  
US-08-728-323A-1  
Sequence 1, Application US/08728323A  
Patent No. 5946676  
GENERAL INFORMATION:  
APPLICANT: Chang, Yuan  
APPLICANT: Bohenzky, Roy A.  
APPLICANT: Russo, James J.  
APPLICANT: Edelman, Isidore S.  
APPLICANT: Moore, Patrick S.  
TITLE OF INVENTION: Immediate Early Protein From Kaposi's  
TITLE OF INVENTION: Sarcoma Associated Herpesvirus, DNA  
TITLE OF INVENTION: Encoding Same And Uses Thereof  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/728,323A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 0575/52268/JPM/MS/SKS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-278-0400  
TELEFAX: 212-391-0525  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3489 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..3489  
US-08-728-323A-1















COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/267,803B  
FILING DATE: 28-JUN-1994  
CLASSIFICATION: A35  
ATTORNEY/AGENT INFORMATION:  
NAME: McCormack, Myra H.  
REGISTRATION NUMBER: 36,602  
REFERENCE/DOCKET NUMBER: 110.00030120  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-305-1217  
TELEFAX: 612-305-1228  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 234 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-08-267-803B-3

[illegible]

```

1      RESULT 9
2      US-08-998-416-915
3      ; Sequence 915, Application US/08998416
4      ; Patent No. 6239264
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Philippsen, Peter
7      ; APPLICANT: Pohlmann, Rainer
8      ; APPLICANT: Steiner, Sabine
9      ; APPLICANT: Mohr, Christine
10     ; APPLICANT: Wendland, Jurgen
11     ; APPLICANT: Knechtel, Philipp
12     ; APPLICANT: Reibischung, Corinne
13     ; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
14     ; TITLE OF INVENTION: AND USES THEREOF
15     ; NUMBER OF SEQUENCES: 1152
16     ; CORRESPONDENCE ADDRESS:
17     ; ADDRESSEE: No. 6239264artis Corporation
18     ; STREET: 3054 Cornwallis Road
19     ; CITY: Research Triangle Park
20     ; STATE: No. 6239264th Carolina
21     ; COUNTRY: USA
22     ; ZIP: 27709
23     ; COMPUTER READABLE FORM:
24     ; MEDIUM TYPE: Floppy disk
25     ; COMPUTER: IBM PC compatible
26     ; OPERATING SYSTEM: PC-DOS/MS-DOS
27     ; SOFTWARE: Patentin Release #1.0, Version #1.30
28     ; CURRENT APPLICATION DATA:

```



















Fri Mar 1 09:11:59 2002

us-09-668-119-1.rni

Page 10

Search completed: February 28, 2002, 19:18:22  
Job time: 5315 sec

.....



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 28, 2002, 17:59:17 ; Search time 171.43 Seconds  
(without alignments)  
8701.779 Million cell updates/sec

Title: US-09-668-119-1

Perfect score: 1740  
Sequence: 1 atgaggaagctggtgtgctc.....cacatctctccagcatga 1740

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_1101:\*

1:	/SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT:*
2:	/SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT:*
3:	/SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT:*
4:	/SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT:*
5:	/SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT:*
6:	/SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT:*
7:	/SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT:*
8:	/SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT:*
9:	/SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT:*
10:	/SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT:*
11:	/SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT:*
12:	/SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT:*
13:	/SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT:*
14:	/SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT:*
15:	/SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT:*
16:	/SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT:*
17:	/SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:*
18:	/SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT:*
19:	/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:*
20:	/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:*
21:	/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:*
22:	/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1725.8	99.2	2556	22	AA159455
2	1647.2	94.7	3438	22	AA161241
3	939.2	54.0	1388	21	AA659039
4	338.6	19.5	386	22	AA65986
5	241.4	13.9	316	22	AA120017
6	241.4	13.9	316	22	AA145215
7	241.4	13.9	316	22	AA105725
8	241.4	13.9	436	22	AA110775
9	241.4	13.9	436	22	AA132033
10	241.4	13.9	436	22	AA100701
11	191.6	11.0	486	22	AA75507

12	181.2	10.4	3489	21	AAA30290
13	181.2	10.4	3489	22	AA682901
14	181.2	10.4	3207	20	AAV73805
15	181.2	10.4	13707	19	AAV19841
16	171.8	9.9	543	13	AAQ23092
17	167.4	9.6	5120	22	AA684677
18	166.4	9.6	14704	13	AAQ20685
19	163.2	9.4	10266	17	AA733007
20	156.8	9.0	567	21	AA29550
21	153.6	8.8	658	21	AA25602
22	149.2	8.6	1159	21	AA69240
23	147.6	8.5	1037	21	AA59242
24	147.6	8.5	1472	21	AA59241
25	140	8.1	397	20	AA89891
26	140	8.0	6794	21	AA45597
27	131.8	7.6	8201	21	AA88864
28	131.6	7.6	629	21	AA708420
29	131.6	7.6	3015	20	AA32027
30	131.6	7.6	3015	22	AA690084
31	130.8	7.5	234	16	AA084832
32	126.6	7.3	6530	14	AA051557
33	126.6	7.3	6530	20	AA22072
34	126.6	7.3	6530	22	AA68948
35	125	7.2	6453	21	AA45239
36	121	7.0	422	21	AA479295
37	120.6	6.9	477	21	AA24307
38	120.4	6.9	799	19	AA55831
39	120	6.9	799	19	AA55831
40	118.8	6.8	1925	20	AA90924
41	118.8	6.8	1926	21	AA50254
42	118.8	6.8	1926	22	AA682902
43	118.8	6.8	2580	21	AA75454
44	118.8	6.8	5452	20	AA90923
45	118.8	6.8	8705	20	AA23778

#### ALIGNMENTS

RESULT 1	
AA159455	AA159455 standard; cDNA: 2556 BP.
AC	AA159455;
XX	
DI	22-OCT-2001 (first entry)
XX	
DE	Human polynucleotide SEQ ID NO 1658.
KW	Human; noctropic; immunosuppressant; cytosolic; gene therapy; cancer;
KW	peripheral nervous system; neuropathy; central nervous system; CNS;
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW	chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW	leukaemia; ss.
OS	Homo sapiens.
XX	
PN	WO200153312-A1.
XX	
PD	26-JUL-2001.
XX	
PF	26-DEC-2000; 2000WO-US34263.
XX	
PR	21-JAN-2000; 2000US-0488725.
PR	25-APR-2000; 2000US-055317.
PR	09-JUL-2000; 2000US-0586042.
PR	19-JUL-2000; 2000US-0620312.
PR	03-AUG-2000; 2000US-0653450.
PR	14-SEP-2000; 2000US-0662191.
PR	19-OCT-2000; 2000US-0693036.
PR	29-NOV-2000; 2000US-0727344.
XX	

Kaposi's sarcoma-a  
Nucleotide sequenc  
KSHV LTR DNA (nucl  
KSHV long unique c  
Antigen tc-7a gene  
DNA sequence of hu  
PKS 741 insert con  
Mouse SRY-related  
HIV codon altered  
cDNA sequence of a  
An EcoRI fragment  
Exons E, C and A o  
Exons D, C, B and  
Spinocebellar at  
cDNA sequence of h  
Human dentin slalo  
Fusarium venenatum  
Human MERT1 relate  
246970 cDNA clone.  
Spinocebellar at  
Loricrin gene. Ho  
Nucleotide sequenc  
Mouse loricrin gen  
DNA encoding angio  
Human lung tumour-  
Probe #15232 used  
Human SCA7 genomic  
Nucleotide sequenc  
Epstein Barr Virus  
Epstein Barr Virus  
EBV tethering prot  
Nucleotide sequenc  
Anti-sense strand  
Vector pshuttle DN



PA (HYSE-) HXSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao Q, Zhou P, Goodrich R, Dimaano RT;  
XX WPI: 2001-442253/47.  
XX P-PSDB: AAM40299.

PT Novel nucleic acids and polypeptides, useful for treating disorders  
XX such as central nervous system injuries -

XX Claim 1: SEQ ID NO 1658; 10078bp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilization of the activities such as: immune system suppression,  
CC activation/inhibition activity, chemotactic/chemokinetic activity, hemostatic  
CC assays for receptor activity, cancer diagnosis and therapy, drug screening,  
CC C.N.S disorders  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.

XX Sequence 2556 BP; 595 A; 884 C; 678 G; 399 T; 0 other;

Query Match 99.2%; Score 1725.8; DB 22; Length 2556;  
Best Local Similarity 99.8%; Pred. No. 3.2e-293;  
Matches 1738; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 atgaggaagagctgtgtgtgtgacacagtaataccacagacagatggaagcattttc 60  
Db 338 atgaggaagagctgtgtgtgtgacacagtaataccacagacagatggaagcattttc 397  
QY 61 ctgaagagcagaagaccgagacgaataccttctctctgagcagagcattatcattt 120  
Db 398 ctgaagagcagaagaccgagacgaataccttctctctgagcagagcattatcattt 457  
QY 121 cgaagacattacaacaaagaatcgaagcttcgaatgagcagacagcattatcattt 180  
Db 458 cgaagacattacaacaaagaatcgaagcttcgaatgagcagacagcattatcattt 517  
QY 181 agcctgagctgcgga>ctgtgcgagggagccgttggaattgacatgctctcggggccg 240  
Db 518 agcctgagctgcgga>ctgtgcgagggagccgttggaattgacatgctctcggggccg 577  
QY 241 ggaagctctctggagcgggagtgagcttctgtgacatgggacagacaaatgtctctca 300  
Db 578 ggaagctctctggagcgggagtgagcttctgtgacatgggacagacaaatgtctctca 637  
QY 301 gggcagcgcgcctccttggacactcggggatggccctcacaagcagatgctgtctcag 360  
Db 638 gggcagcgcgcctccttggacactcggggatggccctcacaagcagatgctgtctcag 697  
QY 361 gcaactcacacagaccagctgacagctcagcaggtggcgtcgcagcagcagcagcaacag 420  
Db 698 gcaactcacacagaccagctgacagctcagcaggtggcgtcgcagcagcagcagcaacag 757  
QY 421 cagcagcttcacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 480  
Db 758 cagcagcttcacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 817  
QY 481 cagcagcttcacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 540  
Db 818 cagcagcttcacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 877

QY 541 cagcagcagctccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 600  
Db 878 cagcagcagctccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 937  
QY 601 aatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 660  
Db 938 aatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 997  
QY 661 caaacaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 720  
Db 998 caaacaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1057  
QY 721 cag 780  
Db 1058 cag 1117  
QY 781 cccag 840  
Db 1118 cccag 1177  
QY 841 cagcctcagctgtgtcagaaacaaacacacacacacacacacacacacacacacacacac 900  
Db 1178 cagcctcagctgtgtcagaaacaaacacacacacacacacacacacacacacacacacac 1237  
QY 901 ttgtgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 960  
Db 1238 ttgtgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1297  
QY 961 ttgtgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1020  
Db 1298 ttgtgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1357  
QY 1021 cag 1080  
Db 1358 cag 1417  
QY 1081 cagaagcagcctccacatgctgtctcgtcagcagcagcagcagcagcagcagcagcagcagcag 1140  
Db 1418 cagaagcagcctccacatgctgtctcgtcagcagcagcagcagcagcagcagcagcagcagcag 1477  
QY 1141 tcatgtccctccctccctccctccctccctccctccctccctccctccctccctccctccctcc 1200  
Db 1478 tcatgtccctccctccctccctccctccctccctccctccctccctccctccctccctccctcc 1537  
QY 1201 tccaagctcagctgtgcccctgcccacatcccaatgagcttctcgtccacagccctccacag 1260  
Db 1538 tccaagctcagctgtgcccctgcccacatcccaatgagcttctcgtccacagccctccacag 1597  
QY 1261 cagccctccacagccagtgacggtgcgagacccacacacacacacacacacacacacacacac 1320  
Db 1598 cagccctccacagccagtgacggtgcgagacccacacacacacacacacacacacacacacac 1657  
QY 1321 ggaactttaaac 1380  
Db 1658 ggaactttaaac 1717  
QY 1381 gctgaagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1440  
Db 1718 gctgaagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1777  
QY 1441 cgcgcag 1500  
Db 1778 cgcgcag 1837  
QY 1501 aagagccttctggaacattctgacagacccctcgaagcgtgtcccttaagaacatttgaa 1560  
Db 1838 aagagccttctggaacattctgacagacccctcgaagcgtgtcccttaagaacatttgaa 1897  
QY 1561 aagtgagatgcaccttggaagaaactcaagaatgacat-gcggtgagccacacccacagc 1619  
Db 1898 aagtgagatgcaccttggaagaaactcaagaatgacat-gcggtgagccacacccacagc 1957







Db	1238	cgagcctcttggtgtcttcacagagccagacgtctctcccttgagacaatgtgttatalaccacacacaa	1297
QY	955	ctgaaattgtctccagctcagctcagatgtgtgttgtaagaaagccccaatgtgagccaggtgtcag	1014
Db	1298	ctgaaattgtctccagctcagctcagatgtgtgtgtgaagaaagccccaatgtgagccaggtgtcag	1357
QY	1015	cagcagcagaaacagcgtgtatacaagacagcccaagctgtcccaagatgtgtgtgtctccagagtcag	1074
Db	1358	cagcagcagaaacagcgtgtatacaagacagcccaagctgtcccaagatgtgtgtgtctccagagtcag	1417
QY	1075	gtcagcagcagagagagagctccccaactgtctctgtcgtctcaacccgggagcagagatgtgaagcc	1134
Db	1418	gtcagcagcagagagagagctccccaactgtctctgtcgtctcaacccgggagcagagatgtgaagcc	1477
QY	1135	cagcagctgtgagtgtccctcccccacagcgtctcccccagacccgggagccagccagctccacag	1194
Db	1478	cagcagctgtgagtgtccctcccccacagcgtctcccccagacccgggagccagccagctccacag	1537
QY	1195	cccaactcccaagctgaagctctgtgacctgtgcccactctcccaagtaagcttctctgcccagccc	1254
Db	1538	cccaactcccaagctgaagctctgtgacctgtgcccactctcccaagtaagcttctctgcccagccc	1597
QY	1255	ttaacgcagagccc-ttccagagagccagatgtacagctgagcgtggagacccacacagaaattaaagtctcc	1313
Db	1598	ttaacgcagagccc-ttccagagagccagatgtacagctgagcgtggagacccacacagaaattaaagtctcc	1657
QY	1314	ctacactgtgagaccttataacaaacactgtgaacccccaagctctgtcaatgagacccagctgtctc	1373
Db	1658	ctacactgtgagaccttataacaaacactgtgaacccccaagctctgtcaatgagacccagctgtctc	1717
QY	1374	cagccagagctgtgagagagagagagatgaacctgtgacaagctgaagcagcgtgtctgaagtaacatcga	1433
Db	1718	cagccagagctgtgagagagagagagatgaacctgtgacaagctgaagcagcgtgtctgaagtaacatcga	1777
QY	1434	gcccctgtgcgcgcatagtatacaaaagatctgaacaaagacagaaacagaaaaaagagacgtgag	1493
Db	1778	gcccctgtgcgcgcatagtatacaaaagatctgaacaaagacagaaacagaaaaaagagacgtgag	1837
QY	1494	taagatgtgaagagcctctctgtgacaattctgtgacaaagcccttgaaagcgtgttccctctgaagac	1553
Db	1838	taagatgtgaagagcctctctgtgacaattctgtgacaaagcccttgaaagcgtgttccctctgaagac	1897
QY	1554	cttgcaaaagtgatgatactgcgcctctgtgagaagaatctaaagataacat-gcggtgtgacacatcc	1612
Db	1898	cttgcaaaagtgatgatactgcgcctctgtgagaagaatctaaagataacat-gcggtgtgacacatcc	1957
QY	1613	cccaagcgcgcgcgcgtgtgcacaaagcccaaaagacgaatgaactatgacagcagcgtctctgtatgc	1672
Db	1958	cccaagcgcgcgcgcgtgtgcacaaagcccaaaagacgaatgaactatgacagcagcgtctctgtatgc	2017
QY	1673	cgctccgtgcacaacat-cgctcactgtgtcttcaacattccctgtatacgcgacatctgtctc	1732
Db	2018	cgctccgtgcacaacat-cgctcactgtgtcttcaacattccctgtatacgcgacatctgtctc	2077
QY	1733	agccatga 1740	
Db	2078	agccatga 2085	
RESULT 3			
AAC59039			
ID AAC59039	standard; cDNA: 1388 BP.		
XX AAC59039:			
XX 02-FEB-2001	(first entry)		
DE Human secreted prote...	coding sequence SEQ ID NO: 60.		
KW Cytosolic; immunosuppressive; neutrotropic; neuroprotective; antiviral;			
KW Antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer;			
KW Vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;			

cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder; neurological disease; infection; human; secreted protein; ss.

Homo sapiens.

WO20005175-A1.

21-SEP-2000.

09-MAR-2000; 2000WO-US06049.

12-MAR-1999; 99US-0124144.

11-JUN-1999; 99US-0138574.

03-DEC-1999; 99US-0168667.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Ruben SM, Komatsuouli G;

WPI; 2000-638175/61.

P-FSD; AAB27609.

Isolated nucleic acid molecule encoding a human secreted protein is used in preventing, treating or ameliorating a medical condition -

Claim 1; Page 372; 428pp; English.

The invention relates to the isolation of genes AAB58990-A59039 encoding 50 human secreted proteins AAB27560-B27609. The genes can be used to generate fusion proteins by linking to the gene for the human immunoglobulin G Fc portion (S001) for increasing the stability of the fusion protein as compared to the human protein only. The genes and proteins are useful for preventing, ameliorating or treating medical conditions, e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer, and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.

Sequence 1388 BP; 364 A; 378 G; 343 C; 297 T; 6 other;

Query Match	54.0%;	Score 939.2;	DB 21;	Length 1388;
Best Local Similarity	98.9%;	Pred. No. 9.9e-156;		
Matches 955;	Conservative 1;	Mismatches 9;	Indels 1;	Gaps 1
Y	1	atggggaagactgtgtgtgcacacagaaatctccagcaagagatgatggagacatgtttt	60	
Db	99	atggggaagactgtgtgtgcacacagaaatctccagcaagagatgatggagacatgtttt	157	
Y	61	ctgaagagccaaagcccgagacaaatctctctctgtggccagcctcatatccattt	120	
Db	158	ctgaagagccaaagcccgagacaaatctctctctctgtggccagcctcatatccattt	217	
Y	121	cgagacattcataacaagaatctcaagcttcgcagtgatctatgaaatgactccag	180	
Db	218	cgagacattcataacaagaatctcaagcttcgcagtgatctatgaaatgactccag	277	
Y	181	agcctgactgagcagactgctgcgggaacgcctggaatggcatgctctcccgggccg	240	
Db	278	agcctgactgagcagactgctgcgggaacgcctggaatggcatgctctcccgggccg	337	
Y	241	ggacaatctctctggcgcgagatggatgacttggatgccatgtgaagccaatgctctcta	300	
Db	338	ggacaatctctctggcgcgagatggatgacttggatgccatgtgaagccaatgctctcta	397	



















































Ratio: 1.587 Gaps: 30  
Percent Similarity: 43.586 Percent Identity: 29.675

alignment\_block:

US-09-668-119-1 x T03454 ..

Align seg 1/1 to: T03454 from: 1 to: 5262

```

1226 CCGCCCTGGGGGGCCGGGACAGCTCTGGGGGGGATGGTAGCTTTGGTGC 275
1227 |||||||:|||||:|||||:
3519 ProProGlnGlyProGlnGlyMetLeuGly..... 3528
276 CATGGACAGCCAAATGTCTCTCAGGGACGCCGCTCTGGAGCTCGG 325
277 :|||:|||||:|||||:|||||:
3529 .ProAlaGlnValAlaValLeuGlnGlnGlnInHisProGlyAlaLeuGlyP 3545
326 GGATGGCCCTCCACAGCATGGCTGTGCTGTACGGCACTCCACAGACC 375
327 :|||:|||||:|||||:|||||:
3545 roGlnGlyProHisArgGlnValLeuMetThrGlnSerArgValLeuSer 3561
376 CAGCTGCAGCTCCAGCAG.....GT 395
377 :|||:|||||:|||||:|||||:
3562 SerProGlnLeuAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 3578
396 GCGCGCTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 445
397 :|||:|||||:|||||:|||||:|||||:
3578 lThrAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 3595
446 CCGCG..... 450
3595 eTrAlaGlyLeuSerHisLeuGlnGlnGlnGlnGlnGlnGlnGlnGln 3611
451 .....CTACAGCAGCA 461
3612 ProGlyLeuSerAlaGlnProMetGlySerLeuGlnGlnGlnGlnGln 3628
462 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 511
463 :|||||:|||||:|||||:|||||:|||||:
3628 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 3642
512 AGCAGCAGTCCAGCAGCAGTAGTGCAG...CAGCAGCAGCAGCAGCAG 558
513 :|||||:|||||:|||||:|||||:|||||:
3642 lGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 3658
559 CAGCAGCAGCAGCAGCAGCAGCAGCAGTAAATTCATCATCAAAATCAGA 608
560 :|||||:|||||:|||||:|||||:|||||:
3659 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 3671
609 ACAGATACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 658
610 :|||||:|||||:|||||:|||||:|||||:
3671 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 3688
659 TCCACACAGCAGCAGCAGCAGCAGCAG..... 690
660 :|||||:|||||:|||||:|||||:
3688 lGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 3704
691 .....CAGCAGCAGCAGCAGCAGCAGCTTGGCA 716
692 :|||||:|||||:|||||:|||||:
3705 GlnSerArgThrLeuLeuSerProGlnGlnGlnGlnGlnGlnGlnGln 3721
717 GCGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 750
718 :|||:|||||:|||||:|||||:
3721 AlauGlyProGlyMetProAlaGlyProLeuGlnHisPheSerSerProG 3738
750 ..... 750
3738 lValAlaLeuGlyProThrLeuLeuThrGlyGlyGlnGlnAsnThrVal 3754
750 ..... 750
3755 AspProAlaValSerSerGlnAlaThrGlnGlyProSerThrHisGlnG 3771
750 ..... 750

```

```

3771 yGlyProLeuAlaIleGlyThrThrProGlnSerMetAlaThrGlnProG 3788
750 ..... 750
3788 lGlyValAlaGlyProSerLeuSerGlyAspSerGlnLeuLeuValGln 3804
751 :|||||:|||||:|||||:|||||:
3805 ProGlnProGlnProGlnProGlnProGlnSerLeuGlnInProProLeuAr 3821
777 TCTGCCCCAGCAGCTGCAGCAG.....ATGCAT..... 804
778 :|||||:|||||:|||||:|||||:
3821 GlnuProGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 3838
804 ..... 804
3838 lGlyGlySerHisGlyGlnGlnGlySerGlySerSerGlnAlaSerSer 3854
804 ..... 804
3855 ValProHisLeuLeuAlaGlnProSerValSerLeuGlyAspGlnProG 3871
805 :|||||:|||||:|||||:|||||:
3871 ySerMetThrGlnAsnLeuLeuGlyProGlnGlnProMetLeuGlyAsp 3888
851 TTGCTCAGAACCCAACTCACAACCTCCCGGCACAGTGCAGACCCAGCT 900
852 :|||||:|||||:|||||:|||||:
3888 roMetGlnAsnAsnThrGlyProGlnProPro.....LysProGlyPro 3902
901 TTGCTGTCCACAGGCCAGCTCCCTGGACAAATGTGTATACCAGC 950
902 :|||||:|||||:|||||:|||||:
3903 ValLeuGlnSerGlyGlnGlyLeuProGly.....ValGlyIleMetPr 3917
951 ACCACTGAATTTGTCCAGCTCCGATG.....GTGGTCAGCAGCAGCC 994
952 :|||:|||||:|||||:|||||:
3917 oThrValAlaGlyGlnLeuArgAlaGlnGlnGlnGlyValLeuAlaGlyAsn 3934
995 CA.....GTGCAGCCCCAGTGCAGCAGCAG..... 1020
996 :|||||:|||||:|||||:|||||:
3934 roGlnLeuArgHisLeuSerProGlnGlnGlnGlnGlnGlnGlnGlnGln 3950
1021 .....CAGACAGCAGTACACAGCAGCTCAGCTCCAGATGATG..... 1059
1022 :|||||:|||||:|||||:|||||:
3951 LeuMetGlnArgGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 3967
1060 :|||||:|||||:|||||:|||||:
3967 oThrGlnLeuProGlyThrGlnThrInHisProLeuGlnGlnGlnGlnGly 3984
1093 :|||||:|||||:|||||:|||||:
3984 yGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 3999
1138 CAGTGCAGTCCCGCCCTCCCGCCAGCCGCGGCAGCAGTGCAGCAGCC 1187
1139 :|||||:|||||:|||||:|||||:
4000 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 4016
1188 TCCACAGCCCAATCCACAGTCCAGCTTCGC..... 1218
1189 :|||||:|||||:|||||:|||||:
4016 a.....ProProGlyAlaLeuSerThrGlyProValLeuGlyProVal 4031
1219 :|||||:|||||:|||||:|||||:
4031 lSerProThrProProSerSerSer.....ProGlnGlnProGlyAspProSerG 4047
1264 CCGTCCAGAGCCAGCAGTGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1313
1265 :|||||:|||||:|||||:|||||:
4047 lInLeuProSerProSerSerGlnLeu.....ProThrGlnAlaGlnLeuPro 4062
1314 CTCACCTGGACCTTTAAACACACCTGTGACCCAGCTGTCTGACAGCC 1363
1315 :|||:|||||:|||||:|||||:

```







1071	GATCCCGGGAGCCACCATCTGGGACAGCTGAGCTGCTCTACTGCTGCT	1022	
43	.....	44	.....CysLysProValC 47
1021	GCTGCTGC.....TGCACCTGGGGCTGCACCTGGGGCTGCTGCACC	981	
47	yscysCysLysProValCys.....CysCysVal	56	
980	ACCATCGAGAGCTGGACAAATTTCAGTGGTGGGTATACAAACATTGG	931	ProVal
57	ProVal	58	
930	TCCAGGAGAGCTTGCCGCTGTGACACCAAGCGCTGGCTCGGACTGTG	881	
59	.....CysSerCysSerSerCysGly.....GlyCysG	68	
880	GCGGAGATTGTGATGGTGTGTTCTTGACCAACTGGAGAGCTGGGGCTGT	831	
68	LysSerSerCys.....GlyLysCysGlySerCys	77	
830	GGTGGCGGCTGTGTTGTTGTGTGTATGATGATCTGCTGGAGCTGGGGG	781	
78	GlySer.....SerCys.....	81	
780	CAGAGCCTGGAGGGCGGAGCTGTGCT.....TGCTGATCGGTGCT	737	
82	.....GlyLysCysGlySerSerCysCys.....	89	
736	GCTGATTTGGTGGCTGGCCCTCCAAACGCTGGCTGGCTGC.....	696	
90	.....LysProValLysCysCysValProVal	98	
695	TGCTGCTGCTGCTGCTGC.....TGCTGTTGC.....	669	
99	CysSerCysSerSerCysGlyGlyCysLysProCysCysCysGlnSer	115	
668	TGTTGTGGAGCGCTGAGCTGTGTTGCTGAGCTGCTGTTGGTGC	620	
115	rcysLysLysProCys.....CysSerSerGlyCysGlySerSerCysG	131	
619	GCTGATCTGTGCTGATTTTGATGATGACAAATTAATTAGATGCTGCTGC	570	
131	LnsSerSerCysCys.....LysProCysCysCys	140	
569	.....TGCTGC.....TGCTGCTGCTGGAGCTGCTGCTGC.....	540	
141	GlnSerSerCysCysLysProCysCysGlnSerSerCysLysSp	157	
539	TGCTGCACACTCT...GCTTGGAACTCCTCTGTCATGAGCACTGTGCTGAG	494	
157	ocysSerSerSerGlyCysGlySerSerCysCysGlnSerSerCys.....	172	
493	CCTGGAACTGC.....TGCTGTTGC.....TGCTCTGCTGCTGC	459	
173	.....CysLysProCysCysGlnSerSerCysCysLysProCys	186	
458	TGCTGTAGCGCGGCTGC.....TGCGTGGCTGGAGACAGTGTGCTG	418	
187	CysCysGlnSerSerCysCysLysProCysCysGlnSerSerCysC	203	
417	TTGCTGCTGCTGCTGACAGCGCACTGCTGAGC.....TGACAGCTGG	374	
373	TCTGTGGA 366		
220	LysCysGly 222		
seq_name: p1r2:B38346			
seq_documentation block:			
ultra-high sulfur keratin 2 - mouse			
(:Species: Mus musculus (house mouse))			



C>Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 24-Sep-1999  
 C:Accession: A38660; B38346  
 R:Wood, L.; Mills, M.; Hetzenbuehler, N.; Vogell, G.  
 J. Biol. Chem. 266, 4024, 1991  
 A:Title: Serine-rich ultra high sulfur protein gene expression in murine hair and skin  
 A:Reference number: A38660; MUID:91154184  
 A:Accession: A38660  
 A:Molecule type: DNA  
 A:Residues: 1-223 <W02>  
 A:Cross-references: GB:M7760; MUD:9200963; PIDN:AAA40107.1; PID:9200964  
 A:Note: This is a correction  
 R:Wood, L.; Mills, M.; Hetzenbuehler, N.; Vogell, G.  
 J. Biol. Chem. 265, 21375-21380, 1990  
 A:Title: Serine-rich ultra high sulfur protein gene expression in murine hair and skin  
 A:Reference number: A38346; MUID:91065960  
 A:Accession: B38346  
 A:Molecule type: DNA  
 A:Residues: 1-21, 'GGCGSGCGCGSGCGSSCGKPVCC', 22-40, 'GSS', 44-45, 'G', 47-48, 'S', 50, 'GSS',  
 <W02>  
 A:Cross-references: GB:M37759; NID:9200961; PIDN:AAA40106.1; PID:9200962  
 A:Note: the sequence reported in this paper has been corrected. See A38660  
 C:Superfamily: ultra-high-sulfur keratin

alignment\_scores:  
 Quality: 427.00 Length: 256  
 Ratio: 3.389 Gaps: 24  
 Percent Similarity: 49.219 Percent Identity: 37.891

alignment\_block:  
 US-09-668-119-1/rev x B38346 ..

Align seg 1/1 to: B38346 from: 1 to: 223

```

1016 TCGTCACCTGGGCTGCATCGGCGCTGCACACCATCGAGCTCG 967
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
3 CysCys.....GlyCysSerGlyGlyCysGlySerCysGlyGly.. 16
966 GACAAATTTCATGTGGTGGTATACACATTGTCACAGAGAGCTT 917
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 .....CysGlySerSerCysCysGlyProValC 26
916 GCGCGTGCACCAAGCGTGGTGGTGGTGGTGGTGGTGGTGGTGG 867
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
26 yscysCysValPro.....ValCysSerCys.....SerSerCysGly 38
866 GGTGGTGTCTGAGCACTGGAGGCTGCTGGGCTGTGTGGCGGCTG 817
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
39 GlyCysGlyGlyGlyCysGlySerCysGlyGlyCysGlyGly.... 53
816 GTGCTGTGTGTGATGATCTGCTGACGCTGCTGGGAGAGCCAGG 767
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
54 .....CysGlySerCysGlyGlyCysLys..... 61
766 GCGAGGCTGTGCTG-TCGATCGTGGTGGTGAATTGGTGGTGGCC 717
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
62 ..GlyGlyCysGlySerCys...GlyGlyCys...LysGlyGlyCysCys 75
716 TCCAAAGCCGCTGCTG.....TGCTGCTGC.....TGCTGCTGCTG 682
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
76 GlnSerSerCysCysLysProCysCysGlnSerSerCysCysLysP 92
681 CTGCTGC.....TGCTGC.....TGCTGT 662
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
92 ocysCysSerSerGlyLysGlySerSerCysCysGlnSerSerCysCysL 109
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
661 GGAGCTGCAGCTGCTGCTATTCGCTGC.....AGCTGCTGT..... 627
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
109 ysprocysCysGlnSerSerCysLysProCysCysSerSerGly 125
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
626 .....TGCTCTGCTGTATCTGCTGATTTGATGATGACAA 589
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
126 CysGlySerSerCysCysGlnSerSerCysCys..... 136

```

```

588 TTAAATTAGATGCTGCTGC.....TGCTGC.....TGCTGCTGCT 554
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
137 .....LysProCysCysGlnSerSerCysLysProCysCysCysG 152
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
553 GGAGCTGCTGCTGC.....TGCTGCACTACT...GCTTGGAACTGCTGC 513
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
152 InSerSerCysCysLysProCysCysSerSerGlyGlySerSerCys 168
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
512 TGCATGGCACTGCTGAGCGCTGCAACTGC.....TGCTGCTGC..... 474
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
169 CysGlnSerSerCys.....CysLysProCysCysCysGlnSe 181
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
473 .....TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 437
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
181 rserCysCysLysProCysCysGlnSerSerCysCysLysProCysC 198
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
436 GCTGCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 387
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
198 yscysGlnSerSerCysCysLysProCysCysGlnSerSerCys... 213
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
386 AGCTGCAGCTGGGCTGCT 369
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
214 ...CysAlaProValCys 218
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
seq_name: p1r2:T44768

```

seq\_documentation\_block:  
 antifreeze glycoprotein AFGP polypeptide precursor [imported] - Boreogadus saida  
 C:Species: Boreogadus saida  
 C>Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000  
 C:Accession: T44768  
 R:Chen, L.; DeVries, A.L.; Cheng, C.H.C.  
 Proc. Natl. Acad. Sci. U.S.A. 94, 3817-3822, 1997  
 A:Title: Convergent evolution of antifreeze glycoproteins in Antarctic notothenioid f  
 A:Reference number: 222834; MUID:97266653  
 A:Accession: T44768  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-507 <CHE>  
 A:Cross-references: EMBL:U43200; NID:92078482; PIDN:AAC60129.1; PID:92078483  
 A:Genetics:  
 A:Introns: 1/3

alignment\_scores:  
 Quality: 399.50 Length: 568  
 Ratio: 1.297 Gaps: 20  
 Percent Similarity: 54.225 Percent Identity: 29.225

alignment\_block:  
 US-09-668-119-1 x T44768 ..

Align seg 1/1 to: T44768 from: 1 to: 507

```

184 CTGACTGGGAGACCTGCTGCGGAGCGGCTGGAATTGGATGCTGCTG 233
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
2 LeuThrGlySerCysLeuLeuGlyProSerCysThrSerSerProArgTy 18
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
234 GGGCCCGGAGACAGTCTCTGGGCGGAGATGGTGTGCTGTCATGGGAC 283
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
18 rgluGlnHisGlnLeu.....ValAla 27
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
284 AGCCAAATGCTCTCTCAGGCGACGCCGCTCTGGAGCTCGGGGATGGCC 333
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
27 rglProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 40
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
334 CCTCAGCAGATGGCTGCTGTCTACGGCAACTCCAGACCCAGCTGCA 383
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
41 .....AlaThrProAlaThr..AlaAla 47
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
384 GCTCCAGAGTGGCGCTGCACAGCA.....GCAGCAACAGC 421
      ||||| ||||| ||||| ||||| ||||| ||||| |||||

```







alignment block:  
US-09-668-119-1 x A56158 ..

Align seg 1/1 to: A56158 from: 1 to: 1893

```

19 GCACACAGTAAATCCACAGAGATATGAGAGCCATGTTTCTGTGAAGGC 68
   |||||
1010 AAlaHisLeuProAsnSerLysSerValProAla.....LeuHisH1 1023
69 CAAGACCCGGGACGAATACCTTTCTCTGAGCCAGGCTCATTTATCCATT 118
   |||
1023 SerHisThrGlySerGlyThrIleSerLeuAla..... 1033
119 TTGCGAGACATTCATTAAGAAATCTCAAGCTCCGTCATGATGATCATTG 168
   |||||
1034 .....AsnSerLysSerArgSerThr..... 1040
169 AATGACATCCAGAGCCATGACTGGCGGACCTGCT.....GGGGAGCCGC 212
   |||||
1041 HisSerLeuHisAsnAsnThrSerIleMetIleGlyIleGlyAlaAla 1057
213 TGGAAATGCGCATG.....CCTCCTCGGGGCCCGGACAGTCTCTGGGCG 256
   |||
1057 aAlaIleGlyIleMetLeuIleGlyIleProAsnGlySerGlnAsnAlaAsnG 1074
257 GGATGGGTAGC.....TTT 270
1074 IAsnGlyIleAsnGlyIleAsnGlyIlePheTyrGlnAsnLeuSerValTyr 1090
271 GGTGCGCATGGGACAGCAGTAAATGCTCTCTCAGGAGCGCCGCTCTGGGAC 320
   |||||
1091 ArgAlaGlnAsnGlnSerGlnProIleLeuAsnGlnArgProIleAla 1107
321 CTGCGGGATGGCCCTACAGCATGGCTGCTGCTCTACAGGACATGCTCAC 370
   |||
1107 aAlaHisAlaIleMetAsnAlaTyrAsnGlySerSerProLeuAlaProG 1124
371 AGACCCAGCTGAGCTACAGAGTGCGGCGCTGCGACAGCAGCAGCAGCAG 420
   |||
1124 IlnGlnIlnProGlnIlnGlnIlnGlnIlnSerProTyrGlnIlnGlnIln 1140
421 CAGCAGTCCAGCAG..... 435
1141 GlnGlnHisMetGlnAlaAsnAlaAsnLeuProProThrArgProValSe 1157
436 .....CAGTAGCAGCGCGGCTACAGCAGCAGCAGCAGCAGCAGCAGC 472
1157 rAlaTyrTyrHisSerIlnGlnIlnSerAlaGlnIlnGlnIlnGlnIlnG 1174
473 AGCAACAGCAGCAGCTTCAGGCTCAGCAGAGTGCATGACAGCAGCAGCTTC 522
   |||||
1174 IlnGlnIlnGlnIln.....GlnHisSerIleGlnIlnGlnIlnPhe 1187
523 .....CAAGCAGTAGTGCAGCAGCAGCAGCAGCAGCT 551
1188 AAlaLeuSerSerGlyIleLeuAsnGlyIlnGlnIlnGlnIlnGlnIln 1204
552 CCAGCAGCAG.....CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 574
   |||||
1204 sGlnHisGlnIlnLeuThrLeuAsnAsnArgThrLysSerGlnIlnAsnIln 1221
575 AGCATCTAAATTAATTT..... 591
1221 IlnHisThrLeuArgMetGlnIlnMetAlaProSerMetProAsnIln 1237
592 .....CATCATCAAAATCAGCAGCAGCAG.....ATACA 617
1238 SerAsnMetTyrHisIlnGlnIlnIlnHisGlnIlnGlnIlnLeuProLeuG 1254
618 CCAGCAGCAGCAGCAGCTGGCAGCA..... 642
1254 nGlnIlnGlnIlnGlnIlnGlnIlnProLeuMetSerSerGlnIlnSer 1271

```

```

642 ..... 642
1271 etGlnAsnValAsnAspPheAlaGlyIleTyrGlnAsnGlySerLeuGln 1287
643 .....ATACACA 650
1288 TyrArgArgSerGlnLeuHisAsnProSerThrLeuTyrGlnIlnGlnIln 1304
651 GGTGCGATCCACAGCAGATGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 700
   |||||
1304 nGlnIlnGlnIlnGlnIlnGlnIlnGlnIlnGlnIlnGlnIlnGlnIlnGln 1321
701 AGCAGCAGCT.....TTGGAGCCCGACAGCA..... 729
1321 IlnGlnAlaSerProAsnPheIleAlaLeuProProLysProLeuGlySer 1337
730 ATTACAGACGCCAGATGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCT.. 777
   |||||
1338 LeuGlnSerPro.....AsnLysProAsnValProProSerThrAlaIle 1352
778 .CTGCCCGACAGCTGACAGATGCATCAGCAGCAG.....C 814
   |||||
1352 oLysProGlnIlnGlnIlnGlnIlnArgTyrIleGlnIlnSerLeuProAlaG 1369
815 ACCACAGCCCGCCACACAGCC..... 837
1369 IlnAspLysProProLeuProProThrAlaThrHisProLeuPheLysAla 1385
838 .....CAGCAG.....CC 845
1386 ThrGlnGlnIlnIleAlaProGlyMetAsnTyrValAlaSerThrLeuAsp 1402
846 TGCA.....GTGCTCAGAACCAACCATCAGCAGCAGCAGCAGCAGCAGCAG 883
   |||||
1402 oProLysGlySerTyrValAlaIleSerAsnGlnIlnGlnIlnAsnAlaArgProLeuH 1419
884 AGTCGACAGCCAGCCT..... 900
1419 IlnSerIlnSerAsnProTyrGlnArgGlnIlnGlnIlnLysAspLeuGln 1435
901 .....TTGCTGTCAAGAGC 914
1436 MetArgArgGlnHisIleArgGlnTyrArgGlnIlnIleSerGlnIln 1452
915 GCAAGCTCTCCCTGACAAATGTTGTATACCCACACACCATGAAA.... 960
   |||||
1452 uSerGlnIlnValIleSerArgSerProMetGlnIlnGlnIlnLeuLysThrL 1469
961 .....TTTGTCCAGCTCGATGGTGGTGCAGCAGCC 993
1469 euIleLeuGlnArgAspPheGlnArgArgAlaGlnIlnLeuGlnIln 1485
994 CCAGTCCAGCCCGAGGTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1043
   |||||
1486 .....GlnGlnIlnAspGlnIlnGlnIlnTyrAspLysGlnAsnValG 1500
1044 GGTGCTCCAGATGATGATGCTCCGAGGTCCAGGTCAGCAGCAGCAGCAGCAGC 1093
   |||||
1500 nGlnLeuPheArgLeuAlaGlyGlyGlyGlnValIleAlaIleGlnIlnP 1517
1094 CCATGCTGCTCTCCGCTCAGCAGCGGCGCAGCAGTGCAGACCCCGCAGTCC 1143
   |||||
1517 rIleThrSer.....TyrArgGlnIlnThrIlnLys 1527
1144 ATGCCCCCTCCCGCCAGCGCTCCCGCAGCGCGGCGCAGCCAGCTCACA 1193
   |||||
1528 LeuAlaGlnMetPro.....As 1533
1194 GCCCAACTCCAAAGTCAAGCTTGGGCTGCGCCCATCTCCCACTAGCTTCC 1243
   |||||
1533 pSerAsnSerLeuValAspSerValPro...ProGlnPro..... 1545
1244 TGCACAGCCCTCAGCGCAGCC.....TCCAGAGCCAGCTG 1281

```



```

1546 ..ProAlaProThrAlaGlnProLeuSerSerAsnThrGlnProIlys 1561
1282 AGGGGGGGACCCACAGAACTTCAGTGTCCCTCCACCTGGACCTTAA 1331
1562 SerIleLeuLysHisAsnArgTyrSerGlnGlyValGlyProSerG1 1578
1332 CACACCTGTGAACCCACGCTGTGTATGATGAGCCAGCTGAGCTCCAGCAGG 1381
1578 yAlaProSerSerProSerLysSerGlnLysSerAlaSerPheAlaAsnG 1595
1382 CTGAGAGACGACGATACCTGGACAACTGAGACACCTGTCCGATACATC 1431
1595 LAArGHIsLeuHisThrGlnHisProIleSerAsnLeuAlaLysGluLeu 1611
1432 GAGCCCTCGCCGCGATGATCAACAAGATCGACAGAAGAA 1473
1612 AsnGlnLeu...ThrMetLeuAspLysAspAsnAsnGln 1624

```

seq\_name: p1r2:T24490

seq\_documentation\_block:

```

hypothetical protein T05A10.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T24490
R:Stuison, J.
submitted to the EMBL Data Library, November 1995
A:Reference number: Z19898
A:Accession: T24490
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2251 <WIL>
A:Cross-references: EMBL:Z68108; P1DN:CA92133.1; GSPDB:GN00028; CESP:T05A10.1
A:Experimental source: clone T05A10
C:Genetics:
A:Gene: CESP:T05A10.1
A:Map position: X
A:Introns: 188/3; 240/3; 420/1; 570/3; 596/1; 732/3; 778/3; 851/3; 1359/2; 1394/2; 1434/

```

```

alignment_scores:
  Quality: 397.00      Length: 599
  Ratio: 1.481        Gaps: 24
  Percent Similarity: 44.741  Percent Identity: 29.549

```

alignment\_block:

US-09-668-119-1 x T24490 ..

Align seg 1/1 to: T24490 from: 1 to: 2251

```

361 GCAACTCCACAGCCAGCTGCTCCAGCGAGTGGCGCTCCAGCAGA 410
111 |||:||||: ||| ||| ||| ||| ||| ||| ||| ||| |||
151 ALaArGHIsLeuHisAsnArgTyrSerGlnGlyValGlyProSerG1 167
411 GGAGCAGACGACGAGCTTCAGCAGCAGCAGAGGCGGCTTACAGCAGC 460
111 |||:||||: ||| ||| ||| ||| ||| ||| ||| ||| |||
167 aGlnAlaArGHIsLeuHisAsnArgTyrSerGlnGlyValGlyProSerG1 184
461 AG.....CAGCAGCAGCAG 474
111 |||:||||: ||| ||| ||| ||| ||| ||| ||| |||
184 LValThrProGlnInPheAlaGlnIleLeuHisMetGlnGlnGlnLeu 200
475 CACAGCAGCAGCTTCAGGCTCAGCAGAGATGCCATGACAGCAGCAGTCCA 524
111 |||:||||: ||| ||| ||| ||| ||| ||| ||| ||| |||
201 GlnGlnGlnGlnInPheGln...GlnGlnGlnLeuGlnGlnGlnGlnLeu 216
525 AGCAGTAGTGCAGCAGCAGCAGCAGCTCCAGCAGCAGCAGCAGCAGC 574
111 |||:||||: ||| ||| ||| ||| ||| ||| ||| ||| |||
216 nGlnGlnGlnGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 233
575 AGCATTAATTAATTCATCAT.....CAAAATCAGCAGCAGATACAG 618
111 |||:||||: ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

233 GlnInLeuGlnGlnValLeuGlnIleSerGlnAlaGlnGlnAlaGln 249
619 CAG.....CAGAACACAGCAGCTGACAGGAAATAGACAGCTGCA 656
250 GlnAlaGlnHisValGlnSerArgGlnMetGlnProSerGlnGlnSerG1 266
657 GCTCCAAACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 706
266 nValGlnAlaGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 283
707 AGGCTTTGGAGGCCACGACCAATTGACGAGCCCGATGACGAGCAGCA 756
283 GlnInLeuSerGlnGlnGlnAlaGlnGlnGlnGlnGlnGlnGlnGln 299
757 CAGCCTCGCCCTCCAGGCTGCTCCAGCAGCAGCTGACGAGATGATCA 806
300 Gln.....LeuGlnGlnInPheLeuGlnGlnGlnGlnGlnHis... 312
807 CACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCTGCAATGTC 856
313 ...GlnGlnArgAlaAlaAlaGlnGlnAlaGln.....AlaG 324
857 AGAACCAACATTCACAACTCCGCGCAGATGCGACAGCCCTTTGGTG 906
324 LAsnAsnAlaSerGlnGlnArgProSerValAlaSerThrProAlaLeu 340
907 TCA.....CAGGCGCAGC 920
341 SerSerThrProGlnLeuAsnAspLeuThrGlnThrMetGlnAlaGln... 356
921 TCTCCCTGGACAAATGTTATACCAACACACCACTGAATTTGTCGAG 970
357 ...LeuGlnGlnGlnLeuLeuGlnGlnGlnGlnGlnAlaGlnAlaGln 973
971 CTCGATGCTGTGTCAGACGCCAGCTGACAGCCCGAGTGTCAGCAGCAG 1020
373 lAglnGln...AlaGlnGlnAlaGlnLeuAlaGlnGlnAlaGlnGln 388
1021 CAG.....ACAGCAGTACAGCAGC 1040
389 GlnGlnGlnGlnSerGlnAsnArgThrValSerGlnAlaLeuGlnIlyr11 405
1041 TCAGGCTGCCAGATGCTG.....GCTCCGAGATGTCAGGTCAGCCAGA 1084
405 eGlnSerMetGlnLeuGlnGlnGlnArgAlaAspGlyThrProAsnAlaGln 422
1085 GCAGCCTCCCATGCTGTCTGCGCGTCCAGCGGCGGCGCAGAGTGCAGACC 1134
422 eGlnGlnGlnArgLeuAlaGlnMetLeuAsnGlnGlnGlnArgMet 438
1135 CCGCAGTCGATGCCCTCCGCCAG..... 1161
439 LeuGlnAsnGlnAlaArgGlnAlaGlnHisArgGlnLeuLeuIleSer 455
1162 ....CGGTCCCGGAGCGCGGCGCAGCCAGCTCAGCCCAACTGCAAG 1207
455 rThrProAlaProArgGlyGlyIleThrMetGlyThrProIleGlyLea 472
1208 TCAGCTCTGGCCCTGCCCATCTCCAGTAGCTTCTGCCAGGCCCTCA 1257
472 LaArgArgGlnGlnGlnProValProSerThrValAlaValThrThrAla 488
1258 CCGCAGCCCTCCAGAGCCAGTGCAGGCGGAGGCCACAGAACTTCAG 1307
489 ProAlaAlaValArgThrProValAlaValProMetLeuLysGlnAsn 505
1308 TGTCCTCCCTACCTGGACCTTAAACACACCTGTAACCCAGCTCT.... 1353
505 rAsnProSerMetAsnProSerSerThrSerThrSerAlaSerAlaThr 522
1354 .....GTCAATGAGCCAGCTGCTCCAGCAGCAGCTGAGAGCAG 1392
522 eSerHisGlnIleLeuAlaProSerLeuSerLysProLeuGlnGlnPro 538

```







```

640 .....CGA 642
887 rgGlyProGlnSerAsnProAsnAlaValProGlyGlyAsnAlaAlaAsn 903
643 ATAGCAGACAGCTGAGCTCCAGACAGACAGACAGACAGACAGACAGACAGACAG 692
904 AlaThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 920
693 CGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 720
920 ngGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 937
720 ..... 720
937 ysgGlnThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 953
720 ..... 720
954 lIeGlnValSerAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 970
721 .....CAGCCAC 727
970 rAsnValSerValAlaAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 987
728 CAATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 777
987 lAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1003
778 CTGCGCCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 816
1004 AsnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1020
817 .....CACCAGCCGACGACGACGACGACGACGACGACGACGACGACGACG 846
1020 ngGlyGlyProAsnGlyProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1037
847 .....CCAGTGTCTGCAAGCAACCA 867
1037 snAsnSerAsnValProSerAspGlyPheSerLeuSerGlnSerGlnSer 1053
868 TCACAACTCCGCGCAGACGACGACGACGACGACGACGACGACGACGACGACG 917
1054 MetAsnThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1070
918 AGCTCTCCCTGGACAAATGTTGATACCAACCAACCAACCAACCAATTTGTC 967
1070 aAlaAlaAlaAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1079
968 GAGCTCCGATGGTGTGACGACGACGACGACGACGACGACGACGACGACGACG 1017
1080 .AlaAlaAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1095
1018 CAGCAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 1067
1096 ArgGlnThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1110
1068 AGTCCAGTCCGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1117
1111 .....AlaAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1125
1118 GC.....CAGCAGGTCGACGACGACGACGACGACGACGACGACGACGACGAC 1146
1125 lGlyGlnValProLeuMetGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1138
1147 CCCCCCTCCCGCCGACGACGACGACGACGACGACGACGACGACGACGACGACG 1196
1139 .....GlyGlyValProValGlyAlaGln 1146
1197 CAATCCAGGTCAGCTCTGCGCCGACGACGACGACGACGACGACGACGACGACG 1246
1146 ySerGlyAsnAlaSerValGly..... 1153

```

```

1247 CCAGCCCCCTACCCGACGACGACGACGACGACGACGACGACGACGACGACGAC 1296
1154 .....ValProValSerAlaGlyPro 1161
1297 CAGAACTTCACT..... 1308
1162 AsnAsnGlyAlaMetAsnGlnLeuGlyGlyProMetGlyGlyMetProGln 1178
1309 .GTCCCCCTACCTGACCTTTAAACACACCTGTGACCCGACGACGACGACGAC 1357
1178 yMetGlnMetGlyGlyProGlyGlyValProLeuAsnProMetGlnMetAla 1195
1358 TGAGCCGACGCTGGCTCCGACGACGACGACGACGACGACGACGACGACGACG 1385
1195 snProAsnGlyGlyAlaProAsnAlaGlnMetMetMetGlyGlyAsnGly 1211
1386 GGAGCAGCAGTACCTGACCAAGCTGACAGCAGCTGTC.....GA 1423
1212 GlyGlyProValProAlaAlaSerGlnAlaLysPheLeuGlnGlnGlnGln 1228
1424 ACTACATGAGCCCTGCGCCGACATGACCAAGATGACAGACAGACAGAA 1473
1228 nIleMetArgAlaGlnAlaMetGlnHisGlnGln..... 1239
1474 GACAGAAAAAGCAGCTGATACATGAAGACCTTCTGACAAATTCGAC 1523
1239 ..... 1239
1524 AGACCCCTCGAAGCGGTGTCCTCGAAGACCTTGCAAAAGTGTGAGATCG 1573
1239 ..... 1239
1574 CCCTGGAGAACTCAAGATGACATGCGGTCGACCCGACCCGACCCGACCC 1623
1240 .....GlnValGlnGlnHisMetAlaGlyAlaArgProProProProG 1254
1624 CGGTGCCACCGACCAAGACGACGACGACGACGACGACGACGACGACGACGAC 1652
1254 lutyAsnAlaThrLysAlaGlnLeuMet 1263

seq_name: p1r2:S25365
seq_documentation_block:
CYC8 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: glucose repression mediator; protein YBR0908; protein YBR112c; SSN
C:Species: Saccharomyces cerevisiae
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 11-Jan-2000
C:Accession: S25365; S48277; S45980; S25404; S25405; A30906; S44692
R:Mannhaupt, G.; Stucka, R.; Ehmele, S.; Vetter, I.; Feldmann, H.
Yeast 8, 397-408, 1992
A:Title: Molecular analysis of yeast chromosome II between CMD1 and LYS2: the excisio
A:Reference number: S25364; MUID:92327848
A:Accession: S25365
A:Molecule type: DNA
A:Residues: 1-966 <MAX>
A:Cross-References: EMBL:X66247; NID:93548; PIDN:CAA46973.1; PID:93550
R:Mannhaupt, G.; Stucka, R.; Ehmele, S.; Vetter, I.; Feldmann, H.
Yeast 10, 1363-1381, 1994
A:Title: Analysis of a 70 kb region on the right arm of yeast chromosome II.
A:Reference number: S48255; MUID:95208357
A:Accession: S48277
A:Molecule type: DNA
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-966 <MAX>
A:Cross-References: EMBL:X78993; NID:9476045; PIDN:CAA55615.1; PID:9476068
R:Feldmann, H.; Mannhaupt, G.; Schwarzlöser, C.; Vetter, I.
submitted to the Protein Sequence Database, August 1994
A:Reference number: S45927
A:Accession: S45980
A:Molecule type: DNA
A:Residues: 1-966 <FE2>
A:Cross-References: EMBL:Z35981; NID:9536449; PIDN:CAA85069.1; PID:9536450; MIPS:YBR1

```



R:Schultz, J.; Carlson, N.  
Mol. Cell. Biol. 7, 3637-3645, 1987  
A:Title: Molecular analysis of SSN6, a gene functionally related to the SNF1 protein kin  
A:Reference number: S25404; M0ID:88065502  
A:Accession: S25404  
A:Molecule type: DNA  
A:Residues: 1-546, 'K', 546-966 <SCH>  
A:Cross-references: EMBL:M17826; NID:g172725; PIDN:AAA35103.1; PID:g172726  
R:Trumbly, R.J.  
Gene 73, 97-111, 1988  
A:Title: Cloning and characterization of the CYC8 gene mediating glucose repression in Y  
A:Reference number: S25405; M0ID:89211964  
A:Accession: S25405  
A:Molecule type: DNA  
A:Residues: 1-546, 'K', 546-966 <TRU>  
A:Cross-references: EMBL:M23440; NID:g171349; PIDN:AAA34545.1; PID:g171350  
A:Gene: SSN6; SSN6; CRT8  
A:Cross-references: SGD:S0000316; MIPS:YBR112c  
A:Map position: 2R  
C:Function:  
A:Description: required for complete derepression of ICUL1; required for repression of SU  
C:Superfamily: unassigned tetratricopeptide repeat proteins; tetratricopeptide repeat  
C:Keywords: nucleus; transcription regulation  
F:224-257/Domain: tetratricopeptide repeat homology <TT1>  
F:262-295/Domain: tetratricopeptide repeat homology <TT2>  
F:296-329/Domain: tetratricopeptide repeat homology <TT3>  
F:330-363/Domain: tetratricopeptide repeat homology <TT4>  
F:365-398/Domain: tetratricopeptide repeat homology <TT5>

alignment\_scores:  
Quality: 386.50 Length: 486  
Ratio: 1.487 Gaps: 16  
Percent Similarity: 53.498 Percent Identity: 29.630

alignment\_block:  
05-09-668-119-1 x S25365 ..

Align seg 1/1 to: S25365 from: 1 to: 966

```

280 GGACAGCCCAATGTCT.....CTCTAGGGAGCCGCTCTGCGACATC 323
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
447 GLYASnProleuAsnTrArgIleSerAlaGlnSerAlaAsnAlaThrAl 463
324 GGGGATGGCCCTCACAGCATGGCTGCTGTCTAGC..... 360
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
463 AsnMetValGlnGlnGlnHisProAlaGlnInThrProIleAsnSer 480
361 .....GCACnCCAGACAGCCAGCTGACGTC 387
|||||:|||||:|||||:|||||:|||||:|||||:
480 eAlaIleThrMetIleTyrSerAsnGlyAlaSerProGlnLeuGlnAlaGlnAla 496
388 CAGCAGAGTGGCCCTGCAGCAGCAGCAGCAGCAGCTTCCAGCACA 437
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
497 GlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAla 513
438 GCAGCAGCGCGCTACAGCAGCAGCAGCAGCAGCAGCAGCAGT 487
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
513 nAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAla 530
488 TCCAGCTTCAG.....CAGAGTCCATG 510
|||||:|||||:|||||:|||||:|||||:|||||:
530 lAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAla 546
511 CAGCAGCAGCTTCCAGCAGTAGTACAGCAGCAGCAGCAGCTCCAGCACA 560
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
547 GlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAla 563
561 GCAGCAGCAGCAGCAGCAGTAAATTAATTCATCATCAAAATCAGCAG 610
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
563 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 580

```

```

611 AGATACAGCAGCAGCAGCAGCAGCTTGCAGCGAATAGCAGAGCTGACGTC 660
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
580 lGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 596
661 CAACAACAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 710
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
597 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 613
711 TTGGAGGCGCCAGCCCAATTCAGCAGCAGCAGCAGCAGCAGCAGCAG 756
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
613 oTyrIleThrGlnProThrValIleGlnAlaHisGlnLeuGlnProPhe 630
757 .....CAGCTCCGCGCTCCAGGCTGCGCC..... 783
630 eThrGlnAlaMetGlnHisProGlnSerSerGlnLeuProGlnGln 646
784 CAGCAGCTGAGCAGATGATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 833
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
647 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 663
834 GCCCAGCAGCT...CCAGTTGCTCAGACCAACCATCAGCA.....C 874
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
663 nAlaGlnAlaProGlnProGlnProGlnIleGlnHisAsnValGlnGlnVal 680
875 TCCCGCCACAGTCCGACAGCCAGCTTGGTG...TCACAGCGCAGCT 921
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
680 eProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 696
922 CTCCCTGCAAAATGTGTATACCCACACACACCTGAATTTGTCCAGC 971
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
697 AlAlaValSerSerSerThrHisThrGlnAsnAsnThrIleSerProAla 713
972 TCCG.....ATGGTGTGACAGCCGCCA.....G 997
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
713 nProThrHisAlaIleProThrGlnAlaProAlaThrGlyIleThrAsnAla 730
998 TGCAGCCCCAGCTGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCT 1047
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
730 lAlaGlnProGlnValLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 746
1048 GCCAGAGTGGTGGCTCCGAGTCCAGTCCAGCAGCAGCAGCAGCTCCCAT 1097
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
747 AsnIleSerValAsnThrAlaThrSerIleGlnGlnGlnGlnGlnGln 763
1098 GCTGTCTGCGCTGACCGGCGCCAGCAGGTGCAGACCCGCGAGTCGATGC 1147
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
763 uValSerAsnGlnSerProAlaValValGlnSerAsnThrAsnAsn.... 778
1148 CCCTCCCGCCAGCAGCTCCCGCAGCCGCGCCAGCCAGCTCAGCAGCC 1197
|||||:|||||:|||||:|||||:|||||:|||||:
779 .....ThrSerGlnGln 782
1198 AACnTCAAGTCACTGCTGCGCCCTGCGCCATTCCTCAGTCTCCGCGC 1247
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
783 GlnIleProValLysAlaAsnSerIlePro.....SerValIleGln 796
1248 CAGCCCTCAGCCGCTCCAGCAGCCAGCAGCAGCAGCAGCAGCAGCAGCAG 1297
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
796 yAlaGlnGlnProGlnProGlnAlaSerProAlaGlnGlnGlnGlnGln 813
1298 AGAAGTCACTGCTCCCTGACCTGACCTTAACACACCTGTGAACCC 1347
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
813 lAlaIleSerValSerProSerThrLysProLeuAsnThrGlnProGlnSer 829
1348 AGCTCTGTCATG.....AGCCAGCTGCTCCAGCCAGCAGCTGA 1385
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
830 SerSerValGlnProThrValSerSerGlnSerSerThrThrIleValAs 846
1386 GAGCAGCAGTACCTGACAGCAGCTGAAGCAGCTGTGCAAGTACATGAGC 1435
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
846 nAspGlnSerThrAlaGlnThrIle...GlnLeuSerThrAlaThrValP 862
1436 CCCT.....GGCGCATGATCACAAGANTGACAGAGCAGAGCAGCAG 1478

```







```

1209 CAGCTGTGGCCCTGGCCCATCT...CCAGTAGCTTCCTGGCCAGCCCT 1255
    :::: ||| :::: ||| :::: |||
585 aalaprolylaleu(ginglucysproproalaleualaaglymetr 602
1256 CACCGAGCCCTCCCGG..... 1272
    :::: ||| :::: |||
602 htleualaproval(glnelythralahsvalvalysglylyprothr 618
1273 .....AGCCAGTGTGGCGCGGAGCCCGACAGACTTC.....AG 1307
    ||||| ||||| ||||| ||||| ||||| |||||
619 Alaserseprovalvalalaglnvalproalalaphetyrmetglnse 635
1308 TGTCCCTCAGCTGTGG/CCTTTAAACACACCTGTGACCCAGCTCTGTCA 1357
    ||||| ||||| ||||| |||||
635 tvalhisluprogllysalaglnthr..... 644
1358 TGAGCCAGCTGGCTCGAGCCAGCTGAGAGCAGCAGTACCTGGACAG 1407
644 ..... 644
1408 CTGAAGCAGCTGTGTG/AGTACATGAGCCCTGGCCGCGATGACACAA 1457
645 .....LeuAlavalysar 649
1458 GATGACAGACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1507
    :::: ||||| :::: ||||| :::: |||
649 glysAlaglnsergluglnuagaspaspleuseralaleualaaser 666
1508 TTCTGACATTTGTGACAGACCCCTGAGAGCGGTGTCCCTGAAGACCTTG 1557
    :::: ||||| :::: ||| :::: |||
666 alleuprothrlysalaserproalalaglnserpro...Lysvalille 681
1558 CAAAAGTGTGAGATCGCCCTGAGAGAACTCAGACATGACATGCGGTGCC 1607
    :::: ||||| :::: |||
682 glnlu.....Lysasnserleu.glylu 690
1608 ACTCC...CCGACCCG.....CCGCGTCCGACCGACCAA 1639
690 ysalaglnprovalalaserleuasnalasnproproasnseraspleu 706
1640 ACAGCAGTA...CCTATGCCAGCGGCTGCT 1666
707 ValAlaleuAlaprothrproserAlapro 716
seq_name: pir2:A53185

```

```

seq_documentation_block:
G-box-binding factor: slime mold (Dictyostelium discoideum)
C.Species: Dictyostelium discoideum
C.Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 24-Sep-1999
C.Accession: A53185
R.Schmidtzler, G.R.; Fischer, W.H.; Firtel, R.A.
Genes Dev. 8, 502-514, 1994
A.Title: Cloning and characterization of the G-box binding factor, an essential component
A.Reference number: A53185; MUID:94170994
A.Accession: A53185
A.Status: preliminary
A.Molecule type: mRNA
A.Residues: 1-708 <SCH>
A.Cross-references: GB:123075; NID:9456561; PID:AAA21021.1; PID:9456562
C.Superfamily: G-box binding factor
C.Keywords: DNA binding; transcription factor; zinc finger

```

```

alignment_scores:
Quality: 383.00 Length: 697
Ratio: 1.272 Gaps: 23
Percent Similarity: 43.185 Percent Identity: 24.390
alignment_block:
US-09-668-119-1 x A53185

```

Align seg 1/1 to: A53185 from: 1 to: 708

```

130 CATTAACAAGAAATCTCAAGCTTCCTGACATGATCTTAATGATGACATGCA 179
    ||||| :::: ||||| ||||| :::: |||
6 HishlglnlglnlsnserSerSerSerSerSerSerSerSerSer 22
180 GAGCTGAGTGGGGGAGCTGTGGCGGAGCCCGCTGAATTTGACATGCTTC 229
    ||||| ||||| :::: ||||| :::: |||
22 nthrileglnlglyseraspleuserasnleleualaleuproleuprol 39
230 CTCGGGCG..... 237
39 euproserillepethrthrtrhlaaglnasnlglnmetasnproleleu 55
238 .....CCGGAGAGTCT...CTGGCGGAGATGGTAGCTTGTGGCAT 278
    ||||| ||||| ||||| ||||| |||||
56 pheprothrothrserSerleuSerleuSerleuSerleuSerleuSerleu 72
279 GGGACAGCCATGTCT.....CTTCAGAGGAGCCGCTCTGGAGACT 322
    :::: ||||| :::: ||||| :::: |||
72 eleupeuproproserSerleuSerleuSerleuSerleuSerleuSerleu 89
323 CCGGAGTGGCCCTCAGACAGATGCTGTGCTCT..... 357
89 spdlglnlntyrproasprmetproasmetvalasplglnlytglelnlehis 105
358 ...ACGGCAACTTCACAGACCCAGCTGACATTCAGAGTGGCGCTGCA 404
    :::: ||||| :::: ||| :::: |||
106 proasnlglnasnprohlsityrasnlytgrlnleuMetRhemetgl 122
405 GCAGCAGCAGCAAGCAGCAGCAGTTCAGCAGCAGCAGCAGCAGCAGCAGTAC 454
    ||||| ||||| ||||| ||||| |||||
122 nglnglnalaglnlglnasnlglnproproglnglnasnlgln.....G 136
455 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTTCAGCAGCAGT 504
    ||||| ||||| ||||| ||||| ||||| |||||
136 lnglnlglnlsnlglnlglnlglnlglnlglnlglnlglnlglnlgln 152
505 GCCATGACAGCAGTTCAGCAGCAGTTCAGCAGCAGCAGCAGCAGTCCA 554
    :::: ||||| :::: ||||| :::: |||
153 metglnlglnlglnlglnlglnlglnlglnlglnlglnlglnlglnlgln 169
555 GCAGCAGCAGCAGCAGCAGCAGCAGCAGTAAATTAAATTGATCATCAAAATC 604
    ||||| ||||| ||||| ||||| |||||
169 sclnglnmetglnlglnlglnlglnlglnlglnlglnlglnlglnlgln 186
605 AGCAACAGATACAG.....CAGCAGCAACAGCAGTTCAGCAGGATA 645
    ||||| ||||| ||||| ||||| |||||
186 euglnglnlglnlsnlglnlglnlglnlglnlglnlglnlglnlglnlgln 202
646 GCACAGCTGCAGCTTCACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 695
    ||||| :::: ||||| ||||| ||||| |||||
203 glnlglnlglnlsnlglnlglnlglnlglnlglnlglnlglnlglnlgln 219
696 GCAGCAGCAGCAGCTTGGAGGCCAGCCAGCAGCAGCAGCAGCAGCAG 745
    ||||| ||||| ||||| ||||| |||||
219 nglnglnglnlglnlsnlglnlsnserglnproglnglnglnlglnlgln 236
746 TGCAGCAGCAGCAGCAGCTCCGCGCTCCAGGCTTCGCCAGCAGCAGTGTAG 795
    ||||| ||||| ||||| ||||| |||||
236 nglnglnglnlgln.....Hsnglnlglnlgln 245
796 CAGATGCATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 845
    ||||| ||||| ||||| ||||| |||||
246 glnlglnlglnlglnlglnlglnlglnlglnlglnlglnlglnlglnlgln 262
846 T..... 846
262 oqlnserleuSerasnSerlglyasnasnasnasnasnasnasnasn 279
846 ..... 846
279 snasnserasnasnasnasnasnasnasnasnasnasnasnasnsernlgln 295

```







```
232 ..... CCGCGGCCG 240
622 roHISglMetGlnGlnGlyLysProMetSerMetAsnArgProPro 638
241 GGACAGCTCTGGGCGGATGGGTAGCTTGTGTCATGGAGCAAGCAAT 290
639 SerGlnMetHisGlnGlnGlyLysProGlnLeuMetAsnArgProse 655
291 GTCTCTCAGAGGCAAGCCGCTCTGGAGCTGGGATG..... 330
655 rSerGlnMetGlyLysGlnProGlnGly..... GlyMetMetSerArgp 670
331 ..... CCGCCCTCAGCAAGCATGGCTGTC 351
670 roProSerGlnProHisTyrLleGlnGlnGlyProSerSerMet..... 684
352 GTCTCTAGGCAACTCCAGACCCAGCTGACCTCCAGCAGGTGGCG... 399
685 ..... AsnThrProGlnGlnGlnMetArgMetGlnGlnGln 699
400 ..... CTGCAGCAGCAGCAGC 415
699 nasAsnGlnAsnMetAsnGlnSerProHisGlnGlnGlnProLeug 716
416 AACAGCAGCAGCTCCAGCAG..... CAG 438
716 LnhISGlnGlnMetGlnGlnGlnAsnAlaProAsnMetAsnMetProGln 732
439 CAGCAGCGCGGCTACAGCAGCAGCAG..... 465
733 GlnGlnLleArgTyrLleGlnGlnGlnGlnProGlnMetLeugGlnGln 749
466 ..... CAGCAGCAGCAACAGCAGCAGCTCCAGC 493
749 rAsnGlyProProProProHisGlnGlnGlnGlnGlnProValArgV 766
494 CTCAGCAGAGTCCATCAGCAGCAGCTCCAGCAGTACGACAGCAG 543
766 alAlaGlnValAlaGlnGlnProGlnAsnMetGlnLleProProArgGln 782
544 CAGCAGCTCCAGCAGCAGCAG..... 564
783 MetThrProGlnGlnSerGlnSerLleHisGlySerMetGlnAsnSerPr 799
565 ..... CAGCAGCAGCAGC 577
799 ogLlnAsnProGlnMetArgMetAsnLeuAlaProProGlnGlnGlnGln 816
578 ATCTAATTAAATTGCAICATCAAAATCAGCAACAGATACAGCAGCA 627
816 Lh..... LhISGlnAsnGlnProProMetGlnGlnGlnAsnPro 828
628 CAGCAGCTCAGCAGATACAGCAGCTGACGCTCCAAACAGCA..... 672
829 GlnSerAspArgGlnMetProGlnMetGlnGlnProGlyArgGlnLleH 845
673 ..... CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCTT 712
845 sProAsnLleLysArgGlnProGlnHisGlnMetGlnGlnAsnArgSerM 862
713 TGGAGGCCAGCCA..... CCAATTCAGCAGCAGCAGC 744
862 etGlnGlnLysProHisAsnProGlnAsnLleProAspGlnMetMetAsn 878
745 ATGCAGACAGCAGCAGCTCCGCGC..... TCCAGAGCTCTGCCCCAGA 788
879 MetGlnGlnMetGlnGlnAspProMetSerGlnGlnProLeugGlnGln 895
1789 GGTGCAGCAGATGATCAGACAGCAGCAGCAGCAGCAGCAGCA..... 831
895 rProGlnProAsnLeuHisLeuGlnHisHisGlnThrProProMetGln 912
832 ..... CAGCGC..... 837
```

---

```
912 etGlnMetMetGlnProValValThrAspAsnPhleGluProAspAsnAsp 928
837 ..... 837
929 GlnThrLleArgTyrLeasngLysLnhISAlaLleLleasngLysMetGlnTyr 945
838 ..... CAGCAGCTCCAGCTG 853
945 rLysLeuValProGlnGlnArgGlnGlnAsnGlnGlnGlnGlnGlnProAsng 962
854 CTCAGACCAACCATCACCACTC..... 876
962 LngLngLysLnhProMetGlnGlnValHisGlnGlnGlyHisGlnGlnGln 978
877 ..... CCGCCAGATCCGAG..... ACCCA 896
979 GlyrPhleGlnGlnGlnHisGlyProProAsnAsnGlnGlnGlnGlnGln 995
897 GCCTTGGTGTACAGCGGCAAGCTCCCTGACAAATGTTGATACC 946
995 nProProMetGlnAsnGlnGlnAsnLleProLleGln..... HisGln 1010
947 AACCAACATGAAATTGTCCAGCTCCGATGCTG..... 981
1010 LngLngLngLnhAsnMetMetAsnGlnProMetGlnSerAsnAsnGln 1026
982 ..... CTGCAGCAGCCCGCAGTGA 1001
1027 AsnMetAlaAsnLleGlnAsnGlnGlnGlnGlnGlnGlnGlnGlnGln 1043
1002 GCCCGAGGTGCAGCAGCAGCAGCAGCTACAGACACTCCAGCTGCGC... 1050
1043 eGlnGlnGlnGlnGlnGlnGln..... LLeProProAlaLysValAlaP 1058
1050 ..... 1050
1058 rOlyrProArgTyrArgLysLysAsnProAlaGlnGlnProProPro 1074
1051 ..... CAGATGTGGTCCCGAGCTCCAGCTCCAGCTCCAGCAGCAGCT 1091
1075 ProLleGlnGlnThrAsnAlaGlnAspLeuAsnSerMetSerGlyLe 1091
1092 CCGCATG..... CTGTCTCT 1105
1091 uProMetGlyAlaLysGlyLysGlyValMetAlaProSerSerLthrT 1108
1106 GCCGTCACCGGCGCAGCAGGTGCAGACCCCGCAGTGAATGCCCTCC 1155
1108 hrAsqThrPhenISGlnGlnGlnGlnAlaProAsnGlyLeuPro..... 1122
1156 CCGCAGCGCTCCCGCAGCGCGCAGCCGACGTCAGCAGCAGCAACTCCAA 1205
1123 ..... ProAsnAlaGlnProMetLleGlnGlnGlnGlnGln 1135
1206 CGTCAGCTTGCGCCCTGCCCATCTCCAGTGAATCTCTCCAGCGCCCT 1255
1135 nLleAlaGlyLysLysProProAsnGlnAsnLysValThrProGlyLysA 1152
1256 CA..... CCGCAGCGCTCCCGCAGCGCA 1278
1152 snValrProLysAsnGlnGlnGlnSerArgProGlnGlnAlaAsnGlyLys 1168
1279 GTGACGGCGCGGAGCCCAAGAACTTCAGTGTCCCTCAGCTGAGCTTT 1328
1169 ValGlnAlaLysSerSerProAspLeuSerGlnProSerThrSerLysG 1185
1329 AAACACACCTGTGAACCCAGC..... TGTGTCA 1357
1185 nArgProProAlaAsnLysLysAlaAspMetThrValLleArgAlaLleV 1202
1358 TGAAGCCAGCTGGCTCCAGCAGCTGAGAGCAGCAGTACTGTGACAAG 1407
```



```

1202 alserHisValylsGlyserArglysmethrProglInuTyrlleu... 1217
1408 CTGAGACGAGCTGTGAAATCATC...GAGCCCTGGCCGATGATCAA 1454
1218 .....TyrValylsGlnGluLeuAsnArgLeuMetAr 1228
1455 CAAGATGACAGACAGACGAGAAAGAGAC.....CTCAGTAAAG 1498
1228 ggluLeuGluLys...AlaAspMetLysThrAspArgArgTyrlleuSer 1244
1499 TGAAGACCTTCTGTGACATTTCTGACAGACCCCTGAGAGCGGTGCCCTG 1548
1244 aAlaGlyLeuLeuAlaLeuIleGlnGlu..... 1253
1549 AAGACCTTGCAAAAGTGTGAGATGCGCTGACAGAACTCAAGATGACAT 1598
1254 .....GlnLysLeuThrIleIleIlePro..... 1262
1599 GCGGTGCCACCTGCCCGCCAGCGC 1620
1263 .ArgIleProMetProAlaArg 1269

seq_name: pir2:T30160
seq_documentation_block:
Hypothetical protein C37A2.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T30160
R:Le, T.T.; Kemp, K.; Scheet, P.
Submitted to the EMBL Data Library, April 1997
A:Description: The sequence of C. elegans cosmid C37A2.
A:Reference number: Z20746
A:Accession: T30160
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1234 <LEU>
A:Cross-references: EMBL:U97194; PDB:AA52447.1; GSPDB:GN00019; CESP:C37A2.2
A:Experimental source: strain Bristol N2; clone C37A2
C:Genetics:
A:Gene: CESP:C37A2.2
A:Map position: 1
A:Introns: 46/1; 124/1; 159/1; 231/1; 508/3; 740/3; 891/1; 931/1; 975/1; 1014/3; 1083/3

alignment_scores:
Quality: 381.00 Length: 666
Ratio: 1.443 Gaps: 31
Percent Similarity: 39.640 Percent Identity: 27.027

alignment_block:
US-09-668-119-1 x T30160 ..
Align seg 1/1 to: T30160 from: 1 to: 1234

211 GCTGAATTGGATGCGCTCCTCGGGCCCGGGA.....CAGTC 248
403 SerGlySerGlnProThrProValProGlyThrProGlnProGlnG 419
249 TCTG.....GGCGGATGGGTAGCTTTGGTCCATGGAGAGC 286
419 nIleThrProGlnProGlySerLeuGlyProMetGlySerLeuGlyPro 436
287 CAATGCTCTCTCAGAGGACCGCTCTGGGACCTGGGGATGGCCCT 336
436 rothAla.....ProProGlySerGlnProMetAsnPro 447
337 CACAGCATGGGTGTGTGTACGGCAACTCCAGAGACCAGTGCAGCT 386
447 ..... 447
387 CCAGCAGGTGGCGCTGACAGACAGCAACAGCAGAGTTCCAGCAGC 436
|||||

```

```

448 .....GlnGlnGlnArgIleGlnGln 455
437 ACCAGCAGCGCGCG..... 450
455 LngLngLngAlaIleProSerAlaSerAsnSerProLeuValAsnLeu 471
451 CTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTTC.....CA 491
472 LeuSerAsnGlnGlnProProGlnGlnGlnGlnGlnGlnGlnGln 488
492 GGCCTCAGCAGAGTGCATGCAGCAGCAGCAGTCCAGCAGTACTGACAGC 541
488 rAlaGlnGlyLeuSerMetGlnGln.....IleAlaIleIleGlnGln 503
542 ACCAGCAGCTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCTAATTGAATTG 591
503 LngLngLng..... 505
592 CATCATCAAAATCAGCAGACAGATACAGCAGCAGCAGCAGCAGCTCAGCG 641
506 HisGlnGlnTyrlleGlnGlnArgIleLeuGlnGlnGlnGlnGlnGln 519
642 AATAGCAGAGCTGCAGCTCCAAACAACAGCAGCAGCAGCAGCAGCAGC 691
520 .....GlnAlaMetMetGlnGlnGlnGlnGlnGlnGln 530
692 ACCAGCAGCAGCAGCAGCT..... 711
530 aGlnGlnGlnGlnGlnGlnAlaProProProThrThrProAsnProGlyHis 546
712 .....TTGGAGCCCGCAGCAGCA..... 729
547 ProGlnGlyPhePheProThrAsnGlnAlaGlyProProGlyThrProG 563
730 .....ATTACAGACCCAGCAGT..... 747
563 YArgProIleProProTyrlaMetGlyGlnProProMetTyrlHisGln 580
747 ..... 747
580 IArgProGlnGlyGlnMetIleGlnArgMetAsnSerTyrlProGlyAsn 596
748 .....CAGCAGCAGCAGCTCCGCTCCAGCT.....CTGCCCA 785
597 AlaGlnGlnPheArgProProProGlnGlnGlnGlnGlnGlnGlnGln 613
786 GCAGCTGCAGCAGATGCATCAGCAGCAGCAGCAGCAGCAGCAGCAGC 835
613 oGlnGlnGlnGln.....GlnProProAlaProProGlnGlnL 626
836 CCCAGCAGCCTCCAGTTGCTCAGAACCAACCATCACACTCCCGCA... 882
626 euGlnGlnProProValGlnAspThrSerAlaValaIaGluProProLys 642
882 ..... 882
643 LysLysLysArgProThrLysLysGlnLysGlnAlaAlaAlaAlaAla 659
883 .....CAGTCCAGACCCAGCCTTTGGTGCACAGCGCCAGCAGCTCCCTG 928
659 aLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 676
929 GACAATGTTGTATACCAACCAACCACTGAATTTGTCGAGCTCG... 975
676 LysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 692
976 .....ATGGTGTGCAGCAGCC..... 993
693 GlnGlnGlnGlnMetValGlnGlnGlnProGlyTyrlProGlyGlnGln 709
993 ..... 993
709 oGlyMetProProProGlnGlnLysAlaPheProProGlyTyrlLeuProGlyG 726

```



```

994 ..... CCAGTGCAGCCCGGAGGTCAGCAGCAGCAGCAGCAGTACAG 1035
726 LVALGlnProthrngnglnnglnlrrpgrnglnnglnngln 742
1036 ACAGCTCAGGCTGCC ..... CAGATGGTGCTCCCGAGTCCAGGT 1076
743 GlnGlnAaGmetAlaMetPheGlnGlnGlnGlnGlnGlnGln 759
1077 CAGCCAGAGCAGCCTCCCATG.....CTGT 1102
759 oGlyLyLProGlnGlnProMetGlyGlnTTPProGlnGlnAaGlnLeuP 776
1103 CCTGCGCCCTACCGGKCCAGCAGGTGACAGCCCGCAGTCCAGT..... 1146
776 rovalProLyLProProGlyAaAaAaSerThrProAaPAlaValAaGln 792
1146 ..... 1146
793 GlnGlnAaProLleProGlyAlaThrMetGlnHisArGleuSerGlyL 809
1147 ..... CCCCCCTCC..... CCCCAGCCGCTCC 1168
809 uPheAlaProProPrCvaLaserSerGlyLyValHisProGlnGlnGln 826
1169 CGAGCCCGGCGAGCCAGCTCAGCAGCCAGCAGCAGTCCAGCTGAGC 1218
826 lnglnGlySerPheHlsarGserAaPserAaLaserValTySerGly 842
1219 CCTGCCCCATCTCCG/GTACGTTCTCGCCAGCCCTCAGCCGAGCCCTC 1268
843 ..... SerHisThrPheGlnGlnGlnGlnGlnGlnGln 853
1269 CCAGAGCCAGCTAGCAGCGCGAGCCAGCAGAACTTACGTGCTCCCTCAC 1318
853 rGlnSerGlnProThraL.....ValPro...P 862
1319 CTGAGACCTTTAAACAACCTGTGAACCCAGCAGCTGTATAGCCAGCT 1368
862 roGlyProGlnGlnAaA.....AaAProGly 870
1369 GAGCTCCAGCCAGGCTCAGAGCAGCAGTACTGAGCAGAGTCCAGAGC 1418
871 GlySerGlyAaPLeGlyGlnLysAlaLleValAaPLeuLeuAaAa 887
1419 GTGGAAGTACTGAGCCCTGCG..... 1442
887 rSer.....GluProLeuAlaAaPLeuGlyAaPLeuGlyAaPLeuG 901
1443 ..... CCGCATATCAACAGATGACAGAGAGAGA 1475
901 LysAaPLeuAaPLeGlnProMetAaPValGlnAaPValGlnThrProSer 917
1476 CAGAAAAAAGAGCCTGAGTAAGATGAAGACCTTCTGAGATCTTGACAG 1525
918 ThrSerAaGlnGly.....GluArGlnGlnAaPLeuAaPLeuAaPLeu 929
1526 A.....CCCTCGAAGCGGTGCTCCCTGAA..... 1550
929 glLeuAlaLaserLleGlnLeuValAaPLeuGlnValAaLaserGlyArG 946
1551 ..... GACCTTGCAGAAAGTGAAGATGCGCTGAGGAA 1583
946 laGlyAlaPheAaAaLleGlnLeuAlaLleGlnGlnAaPLeuAaPLeu 961
1584 ACTCAAAAGATGACATGCGGTGCCAGCTCCCGCAGCCCGCGGTGCCA 1631
962 .....SerValProLleProLeuProAlaLaserPro 972

seq_name: p1r2:T113606
seq_documentation_block:
hypothetical protein 87B1.5 - fruit fly (Drosophila melanogaster)

```

---

```

C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T113606; S23632
R:Murphy, L.; Harris, D.; Barrell, B.
submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: 217668
A:Accession: T113606
A>Status: preliminary; translated from GR/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1589 <MOR>
A:Cross-References: EMBL:Z98269; NID:e1355202; PID:e1251078; PIDN:CA810975.1
R:Decamilli, M.; Cheng, N.; Plerre, D.; Brock, H.W.
Genes Dev. 6, 223-232, 1992
A:Title: The polyhomeotic gene of Drosophila encodes a chromatin protein that shares
A:Reference number: S23632; MUID:92146957
A:Accession: S23632
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1010, 7, 1012-1192, 1, 1194-1274, 1, 1276-1589 <DEC>
A:Cross-References: EMBL:X63672; NID:g11056; PIDN:CA45211.1; PID:g11057
C:Genetics:
A:Gene: FlyBase:ph-P
A:Cross-References: FlyBase:FBgn0004861; FlyBase:FBgn0004860
A:Introns: 12/2; 595/1; 745/2; 1340/1
C:Superfamily: SAM homology
C:Keywords: DNA binding; nucleus
E:74-80, 247-285, 411-450, 494-650, 727-737, 775-955, 1032-1061/Region: glutamine-rich
F:1510-1576/Domain: SAM homology <SAM>

alignment_scores:
Quality: 378.00 Length: 629
Ratio: 1.24 Gaps: 32
Percent Similarity: 48.331 Percent Identity: 29.571

alignment_block:
US-09-668-119-1 x T113606 ..
Align seg 1/1 to: T113606 from: 1 to: 1589

19 GCACACAGTAAATCCAGCAGAGATATGAGAGCCATGTTTCTGAAGC 68
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
674 AaAaSerValSerThrGlnThrAlaGlnAaGlnSerLeuAaL 690
69 CAAGCCCGGAGCAATAC..... 87
111 111 111 111 111 111 111 111 111 111 111 111 111 111
690 alysmetArGlnAaGlnGlnProValArProAlaLeuAlaThrLeuL 707
88 .....CTTCTCTCGTGCC.....AGGCTCATATCCATT 120
707 ysthrGlnLleGlyGlnValAlaGlnAaGlnAaGlnValAaGlyHisLeu 723
121 CGAGCATTCATTAACAAGAAATCTCAGAGTCCGAGTATCCATGA 170
724 ThrThrValGlnGlnGlnGlnGlnAlaThrAaGlnGlnGlnValAa 740
171 TGCACCTCAGAGCCTGAGCTGCGGAGCCTGCGGAGCCTGGAATG 220
740 n.....AlaAlaGlyAaA 745
221 GCATGCTCTCTGGGCGCCCGGAGCAGCTCTGGCGGAGTGGATGCT 270
745 ysmet..... 746
271 GGTGCATGGAGCAGCAATGTCTCTCAGGAGCAGCCGCTCTGGGAC 320
747 .....ValValMetSerThrThrGlnThrProLleThrLeuG 759
321 CTGGGGAATGGCCCTCAGAGATGCTGTGCTGTCTGAGCAGCACTCC 370
759 nasnglyGlnThrLeuHisAlaLleThrAlaAlaGlyValAaPLeuG 776

```







1 . . . ,



OM of: US-09-668-119-1 to: SwissProt\_39:\* out\_format : pfs  
 Date: Feb 28, 2002 7:37 PM

About: Results were produced by the GenCore software, version 4.5,  
 Copyright (c) 1993-2000 Compugen Ltd.

## Command line parameters:

-MODEL=frame+22p.model -DEV=xlp  
 -O=/gn2.1/USPRO\_pool/US09668119/runat\_28022002\_084718\_19436/app\_query.fasta\_1.1824  
 DB=SwissProt\_39 -OFMT=fasta -SUFFIX=rsp -GAPOP=12.000  
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPL=0.000 -LOAPEXT=0.000  
 -OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=Dlosum62  
 -TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct  
 -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
 -NOR=ext -MINLEN=0 -MAXLEN=200000000  
 -USER=US09668119 -ECGN1\_1.42 -NCPU=6 -ICPU=3 -LONGLOG -NO\_XLPXY  
 -WAIT -THREADS=1

## Search information block:

Query: US-09-668-119-1  
 Query length: 1740  
 Database: SwissProt\_39:\*  
 Database sequences: 100059  
 Database length: 36664827  
 Search time (sec): 78.590000

## score\_list:

Sequence	Strd Orig	Zscore	EScore	Len	Documentation
SwissProt_39:SSN6_YEAST +	385.50	271.26	2.7e-08	966	P14922
SwissProt_39:GBE_DICDI +	388.00	271.38	2.7e-08	708	P36417
SwissProt_39:MAN_DROME +	378.00	263.33	3.3e-08	1596	P21519
SwissProt_39:PHF_DROME +	377.50	263.02	3.3e-08	1589	P39769
SwissProt_39:KRUH_HUMAN -	332.50	258.06	5.4e-07	134	P07560
SwissProt_39:AMVH_YEAST +	339.50	238.40	9.6e-07	1367	P08640
SwissProt_39:ANV_NOTCO +	330.00	235.20	2.5e-06	790	P24855
SwissProt_39:GALY_KLUGA +	337.50	232.11	2.9e-06	1008	P32557
SwissProt_39:FSL_DROME +	327.50	228.05	2.4e-06	2038	P13709
SwissProt_39:SNF5_YEAST +	326.50	232.06	3.3e-06	905	P18480
SwissProt_39:GLT4_WHEAT +	317.50	226.47	7.2e-06	838	P08489
SwissProt_39:GLT5_WHEAT +	316.00	225.46	8.2e-06	839	P10388
SwissProt_39:IDRL_HUMAN +	312.50	221.12	1.0e-05	1185	P54259
SwissProt_39:ILP1_ELMTE +	310.00	228.31	1.9e-05	255	P15714
SwissProt_39:VGLX_HSVB +	310.00	221.73	1.4e-05	797	P28968
SwissProt_39:Y192_HUMAN +	310.00	216.07	1.1e-05	2124	P093074
SwissProt_39:GLTA_WHEAT +	308.00	225.04	2.0e-05	356	P10385
SwissProt_39:KRUH_HUMAN -	305.50	227.66	3.1e-05	159	P26371
SwissProt_39:MUOI_HUMAN +	305.50	216.09	1.8e-05	1255	P15641
SwissProt_39:PRO_DROME +	304.50	214.78	1.9e-05	1403	P29617
SwissProt_39:CPO_DROME +	302.50	218.87	2.0e-05	615	P01617
SwissProt_39:ELAV_DROVI +	300.50	217.83	3.5e-05	519	P23241
SwissProt_39:OVO_DROME +	300.50	213.89	2.9e-05	1028	P51521
SwissProt_39:KAPC_DICDI +	296.50	213.87	4.7e-05	648	P34099
SwissProt_39:YKRF_YEAST +	296.00	212.78	4.7e-05	738	P35732
SwissProt_39:IDC1_DROME +	295.50	210.02	4.4e-05	1123	P18169
SwissProt_39:CCP_DROME +	295.00	213.81	5.6e-05	550	P41046
SwissProt_39:CBP_HUMAN +	294.00	204.53	4.1e-05	2442	P10273
SwissProt_39:YKTF3_CAEBL +	284.50	202.11	0.0001	467	P48997
SwissProt_39:YKTF3_CAEBL +	284.50	202.11	0.0001	467	P48997
SwissProt_39:SMF1_HUMAN +	284.50	195.04	9.9e-05	1902	P08497
SwissProt_39:GLT3_WHEAT +	283.50	209.00	0.0001	660	P08488
SwissProt_39:GLT3_WHEAT +	283.00	209.01	0.0002	313	P04727
SwissProt_39:GDA7_WHEAT +	283.00	208.99	0.0002	319	P04725
SwissProt_39:GDA5_WHEAT +	283.00	208.99	0.0002	319	P04725
SwissProt_39:CBP_MOUSE +	283.00	197.16	0.0001	2441	P45481

seq_name	SwissProt_39:SSN6_YEAST	seq_documentation_block
SwissProt_39:DRPL_RAT +	282.00	200.67
SwissProt_39:BRM_DROME +	282.00	198.79
SwissProt_39:TRCU_EBV +	281.00	194.35
SwissProt_39:ZEST_DROVI +	280.50	203.41
SwissProt_39:HDC_DROME +	280.00	199.85
SwissProt_39:SSN6_YEAST	280.00	199.85
SwissProt_39:DRPL_RAT +	282.00	200.67
SwissProt_39:BRM_DROME +	282.00	198.79
SwissProt_39:TRCU_EBV +	281.00	194.35
SwissProt_39:ZEST_DROVI +	280.50	203.41
SwissProt_39:HDC_DROME +	280.00	199.85
SwissProt_39:SSN6_YEAST	280.00	199.85
SwissProt_39:DRPL_RAT +	282.00	200.67
SwissProt_39:BRM_DROME +	282.00	198.79
SwissProt_39:TRCU_EBV +	281.00	194.35
SwissProt_39:ZEST_DROVI +	280.50	203.41
SwissProt_39:HDC_DROME +	280.00	199.85
SwissProt_39:SSN6_YEAST	280.00	199.85
SwissProt_39:DRPL_RAT +	282.00	200.67
SwissProt_39:BRM_DROME +	282.00	198.79
SwissProt_39:TRCU_EBV +	281.00	194.35
SwissProt_39:ZEST_DROVI +	280.50	203.41
SwissProt_39:HDC_DROME +	280.00	199.85
SwissProt_39:SSN6_YEAST	280.00	199.85
SwissProt_39:DRPL_RAT +	282.00	200.67
SwissProt_39:BRM_DROME +	282.00	198.79
SwissProt_39:TRCU_EBV +	281.00	194.35
SwissProt_39:ZEST_DROVI +	280.50	203.41
SwissProt_39:HDC_DROME +	280.00	199.85
SwissProt_39:SSN6_YEAST	280.00	199.85
SwissProt_39:DRPL_RAT +	282.00	200.67
SwissProt_39:BRM_DROME +	282.00	198.79
SwissProt_39:TRCU_EBV +	281.00	194.35
SwissProt_39:ZEST_DROVI +	280.50	203.41
SwissProt_39:HDC_DROME +	280.00	199.85
SwissProt_39:SSN6_YEAST	280.00	199.85
SwissProt_39:DRPL_RAT +	282.00	200.67
SwissProt_39:BRM_DROME +	282.00	198.79
SwissProt_39:TRCU_EBV +	281.00	194.35
SwissProt_39:ZEST_DROVI +	280.50	203.41
SwissProt_39:HDC_DROME +	280.00	199.85
SwissProt_39:SSN6_YEAST	280.00	199.85
SwissProt_39:DRPL_RAT +	282.00	200.67
SwissProt_39:BRM_DROME +	282.00	198.79
SwissProt_39:TRCU_EBV +	281.00	194.35
SwissProt_39:ZEST_DROVI +	280.50	203.41
SwissProt_39:HDC_DROME +	280.00	199.85
SwissProt_39:SSN6_YEAST	280.00	199.85
SwissProt_39:DRPL_RAT +	282.00	200.67
SwissProt_39:BRM_DROME +	282.00	198.79
SwissProt_39:TRCU_EBV +	281.00	194.35
SwissProt_39:ZEST_DROVI +	280.50	203.41
SwissProt_39:HDC_DROME +	280.00	199.85
SwissProt_39:SSN6_YEAST	280.00	199.85
SwissProt_39:DRPL_RAT +	282.00	200.67
SwissProt_39:BRM_DROME +	282.00	198.79
SwissProt_39:TRCU_EBV +	281.00	194.35
SwissProt_39:ZEST_DROVI +	280.50	203.41
SwissProt_39:HDC_DROME +	280.00	199.85
SwissProt_39:SSN6_YEAST	280.00	199.85
SwissProt_39:DRPL_RAT +	282.00	200.67
SwissProt_39:BRM_DROME +	282.00	198.79
SwissProt_39:TRCU_EBV +	281.00	194.35
SwissProt_39:ZEST_DROVI +	280.50	203.41
SwissProt_39:HDC_DROME +	280.00	199.85
SwissProt_39:SSN6_YEAST	280.00	199.85
SwissProt_39:DRPL_RAT +	282.00	200.67
SwissProt_39:BRM_DROME +	282.00	198.79
SwissProt_39:TRCU_EBV +	281.00	194.35
SwissProt_39:ZEST_DROVI +	280.50	203.41
SwissProt_39:HDC_DROME +	280.00	199.85
SwissProt_39:SSN6_YEAST	280.00	199.85
SwissProt_39:DRPL_RAT +	282.00	200.67
SwissProt_39:BRM_DROME +	282.00	198.79
SwissProt_39:TRCU_EBV +	281.00	194.35
SwissProt_39:ZEST_DROVI +	280.50	203.41
SwissProt_39:HDC_DROME +	280.00	199.85
SwissProt_39:SSN6_YEAST	280.00	199.85
SwissProt_39:DRPL_RAT +	282.00	200.67
SwissProt_39:BRM_DROME +	282.00	198.79
SwissProt_39:TRCU_EBV +	281.00	194.35
SwissProt_39:ZEST_DROVI +	280.50	203.41
SwissProt_39:HDC_DROME +	280.00	199.85
SwissProt_39:SSN6_YEAST	280.00	199.85
SwissProt_39:DRPL_RAT +	282.00	200.67
SwissProt_39:BRM_DROME +	282.00	198.79
SwissProt_39:TRCU_EBV +	281.00	194.35
SwissProt_39:ZEST_DROVI +	280.50	203.41
SwissProt_39:HDC_DROME +	280.00	199.85
SwissProt_39:SSN6_YEAST	280.00	199.85
SwissProt_39:DRPL_RAT +	282.00	200.67
SwissProt_39:BRM_DROME +	282.00	198.79
SwissProt_39:TRCU_EBV +	281.00	194.35
SwissProt_39:ZEST_DROVI +	280.50	203.41
SwissProt_39:HDC_DROME +	280.00	199.85
SwissProt_39:SSN6_YEAST	280.00	199.85
SwissProt_39:DRPL_RAT +	282.00	200.67
SwissProt_39:BRM_DROME +	282.00	198.79
SwissProt_39:TRCU_EBV +	281.00	194.35
SwissProt_39:ZEST_DROVI +	280.50	203.41
SwissProt_39:HDC_DROME +	280.00	199.85
SwissProt_39:SSN6_YEAST	280.00	199.85
SwissProt_39:DRPL_RAT +	282.00	200.67
SwissProt_39:BRM_DROME +	282.00	198.79
SwissProt_39:TRCU_EBV +	281.00	194.35
SwissProt_39:ZEST_DROVI +	280.50	203.41
SwissProt_39:HDC_DROME +	280.00	199.85
SwissProt_39:SSN6_YEAST	280.00	199.85
SwissProt_39:DRPL_RAT +	282.00	200.67
SwissProt_39:BRM_DROME +	282.00	198.79
SwissProt_39:TRCU_EBV +	281.00	194.35
SwissProt_39:ZEST_DROVI +	280.50	203.41
SwissProt_39:HDC_DROME +	280.00	199.85
SwissProt_39:SSN6_YEAST	280.00	199.85
SwissProt_39:DRPL_RAT +	282.00	200.67
SwissProt_39:BRM_DROME +	282.00	198.79
SwissProt_39:TRCU_EBV +	281.00	194.35
SwissProt_39:ZEST_DROVI +	280.50	203.41
SwissProt_39:HDC_DROME +	280.00	199.85
SwissProt_39:SSN6_YEAST	280.00	199.85
SwissProt_39:DRPL_RAT +	282.00	200.67
SwissProt_39:BRM_DROME +	282.00	198.79
SwissProt_39:TRCU_EBV +	281.00	194.35
SwissProt_39:ZEST_DROVI +	280.50	203.41
SwissProt_39:HDC_DROME +	280.00	199.85
SwissProt_39:SSN6_YEAST	280.00	199.85
SwissProt_39:DRPL_RAT +	282.00	200.67
SwissProt_39:BRM_DROME +	282.00	198.79
SwissProt_39:TRCU_EBV +	281.00	194.35
SwissProt_39:ZEST_DROVI +	280.50	203.41
SwissProt_39:HDC_DROME +	280.00	199.85
SwissProt_39:SSN6_YEAST	280.00	199.85
SwissProt_39:DRPL_RAT +	282.00	200.67
SwissProt_39:BRM_DROME +	282.00	198.79
SwissProt_39:TRCU_EBV +	281.00	194.35
SwissProt_39:ZEST_DROVI +	280.50	203.41
SwissProt_39:HDC_DROME +	280.00	199.85
SwissProt_39:SSN6_YEAST	280.00	199.85
SwissProt_39:DRPL_RAT +	282.00	200.67
SwissProt_39:BRM_DROME +	282.00	198.79
SwissProt_39:TRCU_EBV +	281.00	194.35
SwissProt_39:ZEST_DROVI +	280.50	203.41
SwissProt_39:HDC_DROME +	280.00	199.85
SwissProt_39:SSN6_YEAST	280.00	199.85
SwissProt_39:DRPL_RAT +	282.00	200.67
SwissProt_39:BRM_DROME +	282.00	198.79
SwissProt_39:TRCU_EBV +	281.00	194.35
SwissProt_39:ZEST_DROVI +	280.50	203.41
SwissProt_39:HDC_DROME +	280.00	199.85
SwissProt_39:SSN6_YEAST	280.00	199.85
SwissProt_39:DRPL_RAT +	282.00	200.67
SwissProt_39:BRM_DROME +	282.00	198.79
SwissProt_39:TRCU_EBV +	281.00	194.35
SwissProt_39:ZEST_DROVI +	280.50	203.41
SwissProt_39:HDC_DROME +	280.00	199.85
SwissProt_39:SSN6_YEAST	280.00	199.85
SwissProt_39:DRPL_RAT +	282.00	200.67
SwissProt_39:BRM_DROME +	282.00	198.79
SwissProt_39:TRCU_EBV +	281.00	194.35
SwissProt_39:ZEST_DROVI +	280.50	203.41
SwissProt_39:HDC_DROME +	280.00	199.85
SwissProt_39:SSN6_YEAST	280.00	199.85
SwissProt_39:DRPL_RAT +	282.00	200.67
SwissProt_39:BRM_DROME +	282.00	198.79
SwissProt_39:TRCU_EBV +	281.00	194.35
SwissProt_39:ZEST_DROVI +	280.50	203.41
SwissProt_39:HDC_DROME +	280.00	199.85
SwissProt_39		



DR SGD; S0000316; CYC8.  
 DR InterPro: IPR001440; TPR.  
 DR Pfam: PF00515; TPR 10.  
 DR SMART: SM00028; TPR; 9.  
 KW Transcription regulation; Repressor; Repeat; TPR repeat;  
 KW Nuclear protein.  
 FT DOMAIN 15 30 POLY-GLN.  
 FT REPEAT 46 79 TPR 1.  
 FT REPEAT 80 113 TPR 2.  
 FT REPEAT 114 147 TPR 3.  
 FT REPEAT 150 183 TPR 4.  
 FT REPEAT 187 220 TPR 5.  
 FT REPEAT 224 257 TPR 6.  
 FT REPEAT 258 291 TPR 7.  
 FT REPEAT 296 329 TPR 8.  
 FT REPEAT 330 363 TPR 9.  
 FT REPEAT 364 398 TPR 10.  
 FT DOMAIN 493 556 30 X 2 AA TANDEM REPEATS OF Q-A.  
 FT DOMAIN 557 587 POLY-GLN.  
 FT CONFLICT 547 547 K -> Q (IN REF. 3).  
 SQ SEQUENCE 966 AA; 107202 MW; 84B509CF3208C5C0 CRC64;

alignment\_scores:  
 Quality: 385.50 Length: 488  
 Ratio: 1.477 Gaps: 17  
 Percent similarity: 53.484 Percent identity: 29.713

alignment\_block:  
 US-09-668-119-1 x SSN6\_YEAST ..

Align seg 1/1 to: SSN6\_YEAST From: 1 to: 966

280 GGACAGCCCATGTCT.....CTTCAGGCGAGCGGCTCTGGAGCTC 323  
 |||||.....:|||||  
 447 GlyAsnProLeuAsnThrArgIleSerAlaInserAlaAsnAlaThrAl 463  
 324 GGGGATGGCCCTCACAGCATGGCTGTGTACG..... 360  
 :|||:.....:|  
 463 AserMetValGlnGlnGlnHisProAlaGlnInThrProIleAsnSer 480  
 361 .....GCACATCGACAGCCGACCTGCAGCTC 387  
 480 eAlaThrMetIySerAsnGlyAlaSerProGlnLeuGlnAlaGlnAla 496  
 388 CAGCAGGTGGCGCTCCAGCAGCAGCAGCAGCAGCTTCACAGCA 437  
 |||||.....:|||||  
 497 GlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGln 513  
 438 GCAGCAGGCGCGCTTCAGCAGCAGCAGCAGCAGCAGCAGCAGT 487  
 |||||.....:|||||  
 513 nAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnAla 530  
 488 TCACAGCTCAG.....CAGAGT 504  
 |||||.....:|||||  
 530 lAglnAlaGlnAlaGlnAlaHisAlaGlnAlaGlnAlaGlnAla 546  
 505 GCCATCGCAGCAGCTTCACAGCAGTACGAGCAGCAGCAGCTCCA 554  
 |||||.....:|||||  
 547 LysAlaGlnAlaGlnAlaGlnAlaGlnAlaGlnGlnGlnGlnGln 563  
 555 GCAGCAGCAGCAGCAGCAGCAGCAGCTCAATTAATGATCATCAATA 604  
 |||||.....:|||||  
 563 nGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 578  
 605 AGCAACAGATACGACGACGACAGCAGCAGCTGCAGCAGATAGCAGCTG 654  
 |||||.....:|||||  
 578 lngInGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 594  
 655 CAGCTCCACACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 704  
 |||||.....:|||||  
 595 GlnLeuGlnGlnLysClyValSerValGlnMetLeuAsnProGlnGln 611

705 GCAGGCTTTGAGGCCCCACCAATTCAGAGCCAGCCAGCTGACAGC 754  
 |||||.....:|||||  
 611 yGlnProTyrIleThrGlnProThrValIleGlnAlaHisGlnLeuGln 628  
 755 CA.....CAGCTCCGCCCTCCAGGCTGTGCC... 783  
 |||||.....:|||||  
 628 rhoSerThrGlnAlaMetGlnHisProGlnSerIleuProPro 644  
 784 .....CAGACGTGACGACGATGATCATCACACAGCAGCAGCCGCC 827  
 |||||.....:|||||  
 645 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 661  
 828 ACCACAGCCCCAGCAGCT...CAGTTCGTCGAACCAACCATCAAA. 873  
 |||||.....:|||||  
 661 nProGlnAlaGlnAlaProGlnProLeuIleGlnHisAsnValGlnGln 678  
 874 .....CTCCGCCACGATGCGACAGCCAGCTTTGGTG...TCACAGCG 915  
 |||||.....:|||||  
 678 snValLeuProGlnLysArgTyrMetGlnGlyAlaIleHisThrLeuVal 694  
 916 CAGCTCTCCCTGGACAATGTGTATACCAACCAACCACTGAATTGT 965  
 :|||:.....:|  
 695 AspAlaAlaValSerSerSerThrHisThrGlnAsnThrLysSerPr 711  
 966 CCGAGCTCCG.....ATGTTGTGACAGACCCCA..... 996  
 |||||.....:|||||  
 711 oArgGlnProThrHisAlaIleProThrGlnAlaProAlaThrGlyIleT 728  
 997 .....GTGCAGCCCGAGGTGACAGCAGCAGCAGCAGCAGTACAGACGT 1041  
 :|||:.....:|  
 728 hAsnAlaGlnProGlnValLysLysGlnLysLeuAsnSerProAsnSer 744  
 1042 CAGGCTGCCACGATGTGCTCCGAGTCCAGGTCCAGCTCAGCAGACGCT 1091  
 :|||:.....:|  
 745 AsnIleAsnLysLeuValAsnThrAlaThrSerIleGlnGlnAlaVal 761  
 1092 CCCCATGTCTCTCCCTGCTCACCGGCGCAGCAGGTGACGCCGCACT 1141  
 :|||:.....:|  
 761 sSerGlnValSerAsnGlnSerProAlaValAlaGlnSerAsnThrAsn 778  
 1142 CGATGGCCCTCCCGCCAGCGCTCCCGACGCGCGCAGCCAGCTCA 1191  
 :|||:.....:|  
 778 sn.....ThrSer 780  
 1192 CAGCCCAATCTCAACTGACACTGTGCGCTGCCCTTCCTCCAGTACTT 1241  
 :|||:.....:|  
 781 GlnGlnGlnLysProValLysAlaAsnSerIlePro.....SerVa 794  
 1242 CTGTCCAGCCCTCACCGCAGCCCTCCAGAGCCAGTGTAGCGCGGA 1291  
 :|||:.....:|  
 794 lIleGlyAlaGlnGlnProGlnGlnAlaSerProAlaGlnGlnAlaT 811  
 1292 CCCACAGAACTTCAGTCCCTCCACCTGAGCTTTAAACACACCTGTG 1341  
 |||||.....:|||||  
 811 hLysAlaAlaSerValSerProSerThrLysProLeuAsnThrGlnPro 827  
 1342 AACCCAGCTGTGTCATG.....AGCCAGCTGCGTCCAGCA 1379  
 :|||:.....:|  
 828 GluSerSerSerValGlnProThrValSerSerGlnSerSerThrThr 844  
 1380 GGCTGAGGACGACGACGACTGACAAAGTGAAGCAGTGTGGAATACA 1429  
 :|||:.....:|  
 844 sAlaAsnAspGlnSerThrAlaGlnThrIle...GluLeuSerThrAlaT 860  
 1430 TCAGAGCCCT.....GGCGGCATGATCAACAATGACGAAGAAGACA 1472  
 :|||:.....:|  
 860 hValProAlaGlnAlaSerProValGlnAspValAlaArgGlnHisSer 876  
 1473 AGACAGAAAAAGGACCTGAGTGAAGACCTTGTGCATTTCTGA 1522  
 :|||:.....:|  
 877 LysGlnGlnLysnGly.....ThrThrGlnAlaSerAlaProSerThr 890  
 1523 CAGACCCCTGGAAGCGGTGTCCTCGAAGACCTTGCAAAAAGTGTGATC 1572



```

      :||| :||| :||| :||| :|||
890 rgluglualagluProAlaAlaSerArgAspAlaLysGlnGlnAspGc 907
1573 GCCCTGGAGAACT 1586
      :||| :|||
907 luthrAlaAlaThr 911

```

seq\_name: SwissProt\_39:GBF\_DICDI

seq\_documentation\_block:

```

ID GBF_DICDI STANDARD: PRT: 708 AA.
AC P36417;
DT 01-JUN-1994 (Rel. 29, created)
DT 01-JUN-1994 (Rel. 29, last sequence update)
DT 01-NOV-1997 (Rel. 35, last annotation update)
DE G-BOX BINDING FACTOR (GBF).
GN GBFA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxId=44689;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=AX3;
RX MEDLINE=94170994; PubMed=8125261;
RA Schmitzler G.R., Fischer W.H., Firtel R.A.;
RT "Cloning and characterization of the G-box binding factor, an
RT essential component of the developmental switch between early and
RT late development in Dictyostelium."
RL Genes Dev. 8:502-514(1994).
CC 1- FUNCTION: CAMP-RESPONSIVE TRANSCRIPTIONAL ACTIVATOR REGULATING
CC LATE GENE EXPRESSION. ESSENTIAL COMPONENT OF THE DEVELOPMENTAL
CC SWITCH BETWEEN EARLY AND LATE DEVELOPMENT. BINDS TO A NUMBER OF
CC CA/GT-RICH GENE REGULATORY ELEMENTS.
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL: L29075; AAA21021.1; -.
DR TRASNFAIC: T00315; -.
KW DictyDdb: DD02046; gbFA.
KW Transcription regulation; Activator; DNA-binding; Nuclear protein;
KW Repeat.
FT REPEAT 339 368 1.
FT REPEAT 481 510 2.
FT DOMAIN 11 21 POLY-SER.
FT DOMAIN 115 263 GIN-RICH.
FT DOMAIN 270 292 POLY-ASN.
FT DOMAIN 549 557 POLY-ASN.
SQ SEQUENCE 708 AA; 79268 MW; B4B6D8F04FACACCA CRC64;

```

alignment\_scores:

Quality:	383.00	Length:	697
Ratio:	1.272	Gaps:	23
Percent Similarity:	43.185	Percent Identity:	24.390

alignment\_block:

us-09-668-119-1 x GBF\_DICDI ..

Align seg 1/1 to: GBF\_DICDI from: 1 to: 708

```

130 CATACAGAAATCTCAAGCTCGTCAGATGATCTATGATGACATCA 179
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
6 HishisglnGlnSerSerSerSerSerSerSerSerSerProserG1 22
180 GAGCCTAGCTGGCGACCTGCTGGCGAGCGCGCTGGAATTGGCATGCCCT 229
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

22 nThrIleGlySerAspLeuSerAsnIleSerAlaLeuProLeuProL 39
230 CTCGGGGGC..... 237
39 euProSerIlePheThrAlaGlnAsnGlnMetAsnProIleLeu 55
238 .....CCGGACAGTCT...CTGGCGGAGTAGGTAGCTTTGGTGCAT 278
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
56 pheProProThrSerSerLeuLeuGlySerSerAsnThrProSerPh 72
GGGACAGCCATGCTCTCTCAGGCGAGCCGCTCTGGAGACT 322
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
72 euLeuProProSerSerIleMetSerSerAsnValPheProSerHisA 89
323 CGGGATGGCCCTTCACAGCATGGCTGCTGCTCT..... 357
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
89 spGlyGlnIleProAspMetProAsnMetValAspGlnIleHis 105
358 ...ACGGCACTCCACAGCCAGCTCCAGCTCCAGAGTGGCGCTGCA 404
: ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
106 ProAsnGlnAsnProHisIleTyrAsnIleGlnIleLeuMetPheMetG1 122
405 GCAGCAGCAGCAGCAGCAGCAGTCCAGCAGCAGCAGCAGCGCGCTAC 454
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
122 nGlnGlnAlaGlnAsnGlnProProGlnGlnAsnGln.....G 136
455 AGCAGCAGCAGCAGCAGCAGCAGCAGTCCAGCTCAGCTCAGCAGAGT 504
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
136 nGlnGlnIleHisIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 152
505 GCCATCAGCAGCAGTCCAGCAGTCCAGTCCAGCAGCAGCAGCAGTCCA 554
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
153 MetGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnH1 169
555 GCAGCAGCAGCAGCAGCAGCAGCAGCAGTCTAATTGATGATCAATAATC 604
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
169 sGlnGlnMetGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 186
605 AGCAGCAGCAGTACAG.....CAGCAGCAGCAGCAGTCCAGCGAGTAA 645
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
186 euGlnGlnIleHisIleGlnIleGlnGlnGlnGlnGlnGlnGlnGln 202
646 GCACAGCTCCAGCTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 695
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
203 GlnGlnGlnGlnHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnH1 219
696 GCAGCAGCAGCAGCTTTGGAGGCCAGCCAGCAATTCAGCAGCAGCAG 745
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
219 nGlnGlnGlnHisGlnHisGlnHisSerGlnProGlnGlnGlnHisA 236
746 TGCAGCAGCAGCAGCTCCGCGCTCCAGAGCTCTCCAGCAGCAGTGGAG 795
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
236 sGlnGlnGlnGln.....HisGlnHisGlnGln 245
796 CAGATGATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 845
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
246 GlnGlnHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnPr 262
846 T..... 846
262 ocIleSerLeuSerAsnSerGlyAsnAsnAsnAsnAsnAsnAsnAsnA 279
846 ..... 846
279 snAsnSerAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnSerHisGln 295
847 .....CAGTTGCTCAGAACCAACATCA..... 870
296 LeuAsnAsnLeuThrLeuSerGlnAsnAsnThrSerGlySerAsnThrPr 312
870 ..... 870
312 oserProSerThrIleGlyLysArgLysHisIleGlnIleThrSerAsnSerG 329

```



```

1
870 ..... 870
329 LulysLysaspserSerGlyGlnThrLleProLysCysThrArgCysAsn 345
870 ..... 870
346 GluaIaIaIaSerTrpLysHisAspLysArgArgTrpTrpCysLysGlyLucy 362
871 .....CAACTCCGCCAC 883
362 sLysLysAlaPheThrProGlyLleThrLysMetGlnGlnIValProGln 379
884 AGTCGACAGACCCAGCTTTGGTGTACAGGCGCAAGCTCTCCGTGACAA 933
379 InlAglInLeuGlnProLeuGlnAsnHisAsnGlnLlePro...Gln 394
934 ATGTGTATACCCACACACCACTGAATTTGTCCGAGCTCCGATGTGTGT 983
395 LeuTrpAspSerGln.....GlnAsnAsnSerSerG 405
984 GCAGCAGCCCGCCAGTCAGCCCGAC..... 1008
405 nasnThrProProthGlnProGlnAsnAsnMetAsnGlnLeuAsnHisG 422
1009 .....GTGACAGCAGTGTG.....CAGACAGCAGTGTGACAG 1035
422 InleuLeuGlnGlnGlnHisGlnGlnInlAglInLeuGlnAlaHisLeuAsn 438
1036 ACAGCTCAGGCTGCCAGATGTGTGCTCCGAGTCCAG..... 1074
439 LeuThrAlaSerAsnGlnGlnIValProProGlnLeuGlnGlnInlLeuAs 455
1075 .....GTGACCCAGAGCAGCCTCC 1093
455 nGlyLysLeuProAsnAsnAsnSerLeuLleThrGlnAsnThrLeuA 472
1094 CCATGCTGCTGCTGCCGTACCCGCGCCAGCAGTGTGACAGCCCGCAGTGTG 1143
472 snSerLeu.....TherValSer 479
1144 ATGCCCTCCCTCCCTCC..... 1158
480 CysProProCysProLeuGlyArgGlyLleSerSerTrpLysHisAsp 496
1158 ..... 1158
496 sLysArgTrpPheCysLysGlyCysLysLysProPheThrProValGlyA 513
1159 .....CAGCCGTCGCCGAGCGCGCCAGCCAGCTCAGCCAGCCAGC 1200
513 lAglLysLeuSerProSerSerProSerProLysLysLysSerAsn 529
1201 .....TCCACGTCAGCTCTGTGCTCCCTGCCCCATC 1229
530 lIleThrProLleThrThrSerSerThrSerSerSerThrProSe 546
1230 T..... 1230
546 rIleIleAsnAsnAsnAsnAsnAsnAlaAsnSerSerLysAsnAsn 563
1231 .....CCGATAGCTCTGCTGCCCCAGCCCTCA 1257
563 hrProLysLysGlnLeuSerProProProSerValLeuGlnSerProSer 579
1258 CCGCAGCGCC.....TCCGAGAGCCAGCTG.....AGGCGCG 1289
580 SerSerSerLleSerGlnSerProLeuGlnLeuAsnTrpGlnThrProth 596
1290 GACCCGACAGACTTCAGTGTCCCTCA...CCTGACCTTTAAACACAC 1336
596 rCysSerProAsnProSerLeuProSerLleGlyLysAsnLeuAsnSerG 613

```

```

1337 CTGNGAAC.....CCGAGCTGTGATGAGCCCGAGTGGCTCCAGC 1377
613 InlAlaAsnSerLleLysProAspGlyLysLeuLleSerGlyLeuSer 629
1378 CAGCTGTGAGGACGACGACGATACCTGACACAGCTCAGCTGTGCAAG... 1425
630 ProProLysSerSerSerSerLeuAsnAsnLeuAsnSerPheSerAsnTh 646
1426 .....TACATGAGCCCTGTGCGCCGCA 1447
646 rGlyAlaLeuLeuSerSerAsnGlyLleAsnLeuAlaAsnLeuGlyAsn 663
1448 TGATCAACAGATGACACAGACGAGAGAGAGAGAGAGAGAGAGAGAGAG 1488
663 rLeuSerGlnLeuAsnLysGlnLysLysArgSerAsp 676

seq_name: SwissProt_39:MAM_DROME

seq_documentation_block:
ID MAM_DROME STANDARD: PRT; 1596 AA.
AC P21519;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-MAY-1992 (Rel. 21, Last annotation update)
DE NEUROGENIC PROTEIN MASTERMIND.
GN MAM.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON-S;
RC MEDLINE=91065516; PubMed=1701150;
RA Smoller D., Friedel C., Schmid A., Bettler D., Lam L.,
RA Vedvobnick B.;
RT "The Drosophila neurogenic locus mastermind encodes a nuclear protein
RT unusually rich in amino acid homopolymers.";
RL Genes Dev. 4:1688-1700(1990).
CC -1- FUNCTION: MAY HAVE A REGULATORY FUNCTION POSSIBLY IN ASSOCIATION
CC WITH THE N GENE PRODUCT.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DEVELOPMENTAL STAGE: DURING EARLY NEUROGENESIS MAM PRODUCTS ARE
CC UBICUOUSLY LOCATED. DURING LATER STAGES THEY ACCUMULATE IN THE
CC CENTRAL NERVOUS SYSTEM.
CC -1- MISCELLANEOUS: THE PROTEIN HAS MANY AA HOMOPOLYMERIC DOMAINS: 21
CC POLY-GLN RUNS (FROM 5 TO 16 AA IN LENGTH), 4 POLY-GLY (6 TO 10
CC AA), 3 POLY-ASN (3 X 5 AA), 1 POLY-ALA (10 AA) AND 1 POLY-THR (5
CC AA) RUNS.
CC -1- SIMILARITY: TO OTHER NUCLEAR PROTEINS OF DROSOPHILA, TO CERTAIN
CC YEAST AND MAMMALIAN REGULATORY PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X54251; CAA38152.1; -.
DR PIR: A33106; A33106.
DR PIR: A36391; A36391.
DR FlyBase: FBgn0002643; mam.
DR Neurogenesis; Nuclear protein; Repeat.
KM DOMAIN 20 84
FT DOMAIN 127 190 GLN-RICH.
FT DOMAIN 196 219 ARG/LYS-RICH (BASIC).
FT DOMAIN 259 304 GLN-RICH.
FT DOMAIN 355 388 ASN-RICH.
FT DOMAIN 392 406 GLY/ASN-RICH.
FT DOMAIN 407 440 GLN-RICH.
FT DOMAIN 407 440 GLY-RICH.

```























```

768 erSerSerThrGluSerSerSerAlaProValProThrProSerSer 784
1583 AACTCAGATGACATCGGTCGCCACGCCGCCGCCGCCGCCGCCGCC 1632
      ::::: ::::: ::::: ::::: :::::
785 SerThrThrGluSerSerSerAlaProValProThrProSerSerThr 801
1633 CGACCAACACAGACGTA...CCTATGCCA..... 1657
      | ::::: ::::: ::::: :::::
801 ThrGluSerSerValAlaProValProThrProSerSerSerAsnI 818
1658 .....GCCGCTCGTGGATGCCGCTCGTGGATGCCGCTCGT 1699
      ::::: ::::: ::::: :::::
818 IeThrSerSerSerAlaProSerSerThrProPheSerSerSer...ThrIu 833
1700 CTTCAACCATTCCTCGTGGATGCCGACATTCGTCGCCGC 1735
      ::::: ::::: ::::: :::::
834 SerSerSerValProValProThrProSerSerSer 845

```

seq\_name: SwissProt\_39:ANP\_NOTCO

seq\_documentation\_block: ANP\_NOTCO STANDARD; PRT; 790 AA.

```

AC P24856;
DT 01-MAR-1992 (Rel. 21, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ANTI-FREEZE GLYCOPOLYMER POLYPROTEIN PRECURSOR (AFGP POLYPROTEIN)
DE [CONTAINS: AFGP7 (AFGP 7); AFGP8 (AFGP 8)] (FRAGMENT).
GN AFGP8.
OS Notothenia coriiceps neglecta (black rockcod) (Yellowbelly rockcod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Notoleosteii;
OC Acanthomorpha; Acanthopterygii; Perciformes;
OC Nototheniidae; Nototheniidae; Notothenia.
OX NCBI_TaxID=8209;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Testis;
RX MEDLINE=91067687; PubMed=2251271;
RA Hsiao K.-C.; Cheng C.-H.C.; Fernandes I.E.; Detrich H.W. III,
RA Devries A.L.;
RT "An antifreeze glycoprotein gene from the antarctic cod Notothenia
RT coriiceps neglecta encodes a polypeptide of high peptide copy
RT number."
RT Proc. Natl. Acad. Sci. U.S.A. 87:9265-9269(1990).
RN [2]
RP REVISIONS TO N-TERMINUS AND 457.
RA Cheng C.-H.C.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ANTI-FREEZE PROTEINS LOWER THE BLOOD FREEZING POINT.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- TISSUE SPECIFICITY: SYNTHESIZED BY THE LIVER AND SECRETED INTO
CC THE BLOOD FROM WHICH THEY BECOME DISTRIBUTED TO ALMOST THE ENTIRE
CC EXTRACELLULAR SPACE.
CC -1- DOMAIN: CONTAINS 44 COPIES OF AFGP8 AND TWO COPIES OF AFGP7.
CC -1- PTM: THE DISACCHARIDE GALACTOSE-N-ACETYLGLACTOSAMINE IS ATTACHED
CC TO THREONINES IN AFGP8 AND AFGP7.
CC -1- DATABASE: NAME=Protein Spotlights;
CC NOTE=Issue 5 of December 2000;
CC WWW=http://www.expasy.org/spotlights/articles/spl1005.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC -----
CC EMBL; M55000; AAA49392.2; -
CC PIR; A38420; A38420.
CC HSSP; P04002; IWFA.
CC InterPro; IPR00104; Antifreeze_1.

```

```

DR InterPro; IPR001778; POA_allergen.
KW Antifreeze protein; Glycoprotein; Polypeptide; Repeat;
KW Multigene family.
FT NON_TER 1
FT PROPEP 1
FT CHAIN 6
FT PEPTIDE 6
FT PEPTIDE 23
FT PEPTIDE 40
FT PEPTIDE 57
FT PEPTIDE 74
FT PEPTIDE 91
FT PEPTIDE 108
FT PEPTIDE 125
FT PEPTIDE 142
FT PEPTIDE 159
FT PEPTIDE 176
FT PEPTIDE 193
FT PEPTIDE 210
FT PEPTIDE 227
FT PEPTIDE 244
FT PEPTIDE 261
FT PEPTIDE 278
FT PEPTIDE 295
FT PEPTIDE 312
FT PEPTIDE 329
FT PEPTIDE 346
FT PEPTIDE 363
FT PEPTIDE 380
FT PEPTIDE 397
FT PEPTIDE 414
FT PEPTIDE 431
FT PEPTIDE 448
FT PEPTIDE 465
FT PEPTIDE 482
FT PEPTIDE 499
FT PEPTIDE 516
FT PEPTIDE 533
FT PEPTIDE 550
FT PEPTIDE 567
FT PEPTIDE 584
FT PEPTIDE 597
FT PEPTIDE 614
FT PEPTIDE 631
FT PEPTIDE 648
FT PEPTIDE 665
FT PEPTIDE 682
FT PEPTIDE 699
FT PEPTIDE 716
FT PEPTIDE 736
FT PEPTIDE 756
FT PEPTIDE 773
FT PEPTIDE 790
SO SEQUENCE 790 AA; 71265 MW; 4CYCCACD48FE902 CMC64;

```

alignment\_scores: Quality: 330.00 Length: 398  
Ratio: 1.460 Gaps: 13  
Percent Similarity: 56.784 Percent Identity: 33.166

alignment\_block:

US-09-668-119-1 x ANP\_NOTCO ..

```

Align seg 1/1 to: ANP_NOTCO from: 1 to: 790
285 GCCAATGCTCTCTCGAGGCGCGCTCGGACCTCGGGGATGCC 334
      |||:|||||:|||||:|||||:|||||:
220 AlaThrAlaAlaLeuAsnPhaAlaAlaThrAlaAlaThrPr 236
      |||:|||||:|||||:|||||:|||||:
335 CTCACACATGCTCTCTCGAGGCGCGCTCGGACCTCGGGGATGCC 377
      |||:|||||:|||||:|||||:|||||:
236 AlaThrAlaAlaLeuAsnPhaAlaAlaThrAlaAlaThrPr 253
      |||:|||||:|||||:|||||:|||||:

```



```

378 .....GCTGCAGCT 386
253 lAlaIatThrProAlaLeuIllePheAlaIatThrAlaIatThrAlaIatThr 269
387 CCAGCAGGTGGCGCTGCAGCAGCAGCAGCAGCAGCAGTCCAGCAGC 436
270 ProAlaIatThrAlaIaCysAsnPheAlaIatThrAlaIatThrProAlaIatH 286
437 AGCAGAGCGCGCGCTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 486
286 rAlaIatThrProAlaLeuIllePheAlaIatThrAlaIatThrAlaIat 303
487 TTCAGGCTGCAGCAGTGCAT.....GCAGCAGCAGTCCAGCAGC 527
303 hrProAlaIatThrAlaLeuAsnPheAlaIatThrAlaIatThrProAla 319
528 AGTAGTCGA.....GCAGCAGCAGCAGTCCAGCAGC 559
320 ThrAlaIatThrProAlaLeuIllePheAlaIatThrAlaIatThrAlaI 336
560 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 609
336 atThrProAlaIatThrAlaLeuAsnPheAlaIatThr.....AlaIat 351
610 CAGATACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 647
351 hrProAlaIatThrAlaIatThrProAlaLeuIllePheAlaIatThrAlaI 367
648 ACAGCTGCAGCTCCAAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 697
368 ThrAlaIatThrProAlaIatThrAlaLeuAsnPheAlaIatThrAlaI 384
698 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 747
384 atThrAlaIatThrProAlaIatThrProAlaLeuAsnPheAlaIatThrAlaI 401
748 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 797
401 lathrProAlaIatThrAlaIatThrProAlaLeuIllePheAlaIa..... 415
798 GATGCATCACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 847
416 .....ThrAlaIatThrAlaIatThrProAlaIatThrAlaIat 428
848 CAGTGTCTGAGAACCAACCACTCCAGCAGCAGCAGCAGCAGCAGCAGCAG 897
428 uAsnPheAla...AlaThrAlaIatThrProAlaIatThrAlaIatThrPro 444
898 CC.....TTTGTGTGCAGAGC.....GCAAGCTCTCCCTGGACAAAT 935
444 lAlaLeuIllePheAlaIatThrAlaIatThrAlaIatThrProAlaIatH 460
936 GTTGTATACCAACCACTGAAATTTGTCGAGCTCCAGTGGTGTGC 985
461 AlAlaLeuIllePheAlaIatThrAlaIatThrAlaIatThrProAlaIatH 477
986 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1035
477 aAlaLeuAsnPheAlaIatThrAlaIatThrProAlaIatThrAlaIatH 493
1036 ACAGCTCAGGCTGCCAGATGGTGTCCGAGTCCAGTCCAGTCCAGCAGAG 1085
493 ..... 493
1086 CAGCCTCCCATGCTGCTCCGCTCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1135
494 .....ProAlaLeuIllePheAlaIatThrAlaIatThrAlaIatThrPro 508
1136 CGCAGTGCATGCTCCCTCCCGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1185
508 oAlaIatThrAlaIaPheAsnPheAlaIatThrAlaIatThrAlaIatThrP 525
1186 AGCTCAGAGCCCAACTCCAGTCAAGTCTGTGCTGCCCTGCCCATCTCCAG 1235

```

```

525 roAlaIatThrAlaIaLeuAsnPheAlaIa..... 534
1236 TAGCTTCCTGCCAGCCCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1285
535 ...ThrAlaIatThrProAlaIatThrAlaIatThrProAlaLeuIllePhe 550
1286 CGCGAGCCCGCAGACTTCAGTGTCCCTCAGCAGCAGCAGCAGCAGCAG 1332
550 lAlaIatThrAlaIatThrAlaIatThrProAlaIatThrAlaIaLeuAsn 566
1333 .....ACACTGTGAACCCAGCTGTCTCATGAGCCAGCT 1368
567 AlAlaIatThrAlaIatThrProAlaIatThrAlaIatThrProAla 580
seq_name: SwissProt_39:GALY_KLUUA

```

```

seq_documentation_block:
ID GALY_KLUUA STANDARD; PRT; 1008 AA.
AC P32257;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE TRANSCRIPTION REGULATORY PROTEIN GAL11.
GN GAL11.
OS Kluyveromyces fragilis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_Taxid=28985;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9202026; PubMed=1923818;
RA Dickson R.C., Hopper J., Mylin L.M., Gerardot C.J.;
RT "Sequence conservation in the Saccharomycetes and Kluyveromyces GAL11
transcription activators suggests functional domains."
RL Nucleic Acids Res. 19:5345-5350(1991).
CC -!- FUNCTION: AUXILIARY TRANSCRIPTION ACTIVATOR FOR GENES ENCODING
GALACTOSE-METABOLIZING ENZYMES. ESSENTIAL FOR NORMAL GROWTH ON
NONFERMENTABLE CARBON SOURCES, FOR SPOULATION AND MATING.
CC MISCELLANEOUS: GAL11 LACKS A DNA-DOMAIN, IT PROBABLY COMPLEXES
WITH GAL4 THAT HAS THE CAPACITY TO BIND DNA. ASSOCIATION BETWEEN
GAL11 AND GAL4 MAY SERVE TO EXPEDITE PHOSPHORYLATION OF GAL4.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M68870; AAA35254.1;
KW Transcription regulation; Activator; Galactose metabolism.
FT DOMAIN 75 84 POLY-GLN.
FT DOMAIN 257 287 GLN-RICH.
FT DOMAIN 364 411 GLN-RICH.
FT DOMAIN 456 464 POLY-GLN.
FT SEQUENCE 1008 AA; 114831 MW; 436D1EBAEA17F7DB CRC64;

```

alignment\_scores: Quality: 327.50 Length: 686  
Ratio: 1.157 Gaps: 24  
Percent Similarity: 41.254 Percent Identity: 23.178

alignment\_block:

us-09-668-119-1 x GALY\_KLUUA ..

Align seg 1/1 to: GALY\_KLUUA from: 1 to: 1008

```

289 ATGTCTCTCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 338
||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
234 MetAsnThrSerValGlnGlnPro.....GlyMetGlyValGln 246

```







```

      829 pleuValHisLysThrProGlnProIleuAspGlnIleAsnGlyThrG 846
1532 CGAAGCGGTGTCCTCCGAGACCTTGCAAAAGT..... 1566
      846 IyLysLysLysLysSerLysAlaIaGlnLysAlaArgAspGlnAspPro 862
1567 ...GAGATCCGCTG.....CA 1580
      863 ValGlnIleSerValArgAsnAsnLysLeuLeuMetProSerLysSerG 879
1581 GAACATCAGATACATGCGGTGCGCCACTCCCGACCGCGCGCGGTGCC 1630
      879 uLysThrLeuArgSerPheLysIleProIleAlaAspIleThrAlaCysP 896
1631 ACGGACCA 1638
      896 heLysPro 898

seq_name: SwissProt_39:FSH_DROME

seq_documentation_block:
ID FSH_DROME STANDARD: PRT: 2038 AA.
AC P13709; P13710;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE FEMALE STERILE HOMEOTIC PROTEIN (FRAGILE-CHORION MEMBRANE PROTEIN).
GN FSH(1)H OR FSH.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89276730; PubMed=2567251;
RA Haynes S.R., Mozer B.A., Bhatia-Dey N., David I.B.;
RT "The Drosophila fsh locus, a maternal effect homeotic gene, encodes
RT apparent membrane proteins."
RL Dev. Biol. 134:246-257(1989).
CC -1- FUNCTION: REQUIRED MATERNALLY FOR PROPER EXPRESSION OF OTHER
CC HOMEOTIC GENES INVOLVED IN PATTERN FORMATION, SUCH AS UBX.
CC -1- SIMILARITY: HIGH, TO HUMAN RING3 PROTEIN.
CC -1- SIMILARITY: CONTAINS 2 BROMODOMAINS.
CC -1- SIMILARITY: CONTAINS 1 ET DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M23221; AAA28540.1; -
DR EMBL: M23222; AAA28541.1; ALT_TERM.
DR EMBL: M15762; AAA70424.1; -
DR EMBL: M15763; AAA70423.1; -
DR EMBL: M15764; AAA70422.1; -
DR PIR: A43742; AAA742.
DR HSSP: P04002; IWEA.
DR FLYBASE: FBgn0004656; fsh(1)h.
DR InterPro: IPR001487; Bromodomain.
DR Pfam: PF00439; bromodomain_2.
DR PRINTS: PR00503; bromodomain_2.
DR SMART: SM00297; BROMO_2.
DR PROSITE: PS00633; BROMODOMAIN_1; 2.
DR PROSITE: PS50014; BROMODOMAIN_2; 2.
KW Developmental protein; Bromodomain; Transmembrane; Repeat.
FT DOMAIN 51 123 BROMODOMAIN 1.
FT DOMAIN 495 567 BROMODOMAIN 2.
FT DOMAIN 945 1106 ET DOMAIN.

```

```

FT TRANSMEM 330 350 POTENTIAL.
FT TRANSMEM 451 471 POTENTIAL.
FT TRANSMEM 750 770 POTENTIAL.
FT TRANSMEM 790 810 POTENTIAL.
FT TRANSMEM 816 830 POTENTIAL.
FT TRANSMEM 874 894 POTENTIAL.
FT TRANSMEM 1731 1751 POTENTIAL.
FT TRANSMEM 1939 1959 POTENTIAL.
FT VARIANT 909 909 G -> A.
FT VARIANT 1022 1022 H -> RPKY.
SQ SEQUENCE 2038 AA; 205332 MW; 849E0706D50A0098 CRC64;

alignment_scores:
Quality: 327.50 Length: 559
Ratio: 1.269 Gaps: 20
Percent Similarity: 46.154 Percent Identity: 26.834

alignment_block:
US-09-668-119-1 x FSH_DROME ..

Align seg 1/1 to: FSH_DROME from: 1 to: 2038

130 CATACAGAAA.....TCACAGCTCCGTCAGTATCC 164
|||||.....|
1226 HISAsnLysAsnGlyProAsnAspLeuSerValGlnProGlyGlyPr 1242
|||.....|
165 TATGATGACCTC.....CAGAGCTGACGCGGAGCTGTGCGG 205
|||.....|
1242 oLksnAlaIaLeuProProHHisSerPheAlaGlyThrAlaThrV 1259
|||.....|
206 GAGCC..... 210
|||
1259 aLAlaThrSerGlnSerSerGlyGlyLeuArgIleAlaSerAsnLeuH 1275
|||
211 .....GCTGGAATTCGATGCTCTCCGCGGCGGAGACTGTGCGG 254
|||.....|
1276 LysProSerGlyLeuGly.....GlyGlyAspLeuG 1286
|||
255 C.....GGGATGGGTAGCT 268
|
1286 yGlnHisHisAlaIaLeuAlaIaIaLeuThrSerGlyLeuAsnSer 1303
|||.....|
269 TTGGTCCATGGAGACAGCCATGTCTCTCAGG..... 303
|||.....|
1303 hrgLysThrAlaGlyGlyGlyLeuAsnAsnGlnGlySerAsnAsn 1319
|||.....|
304 ...CAGCCGCTCTGGGACCTCGGGATGGCCCTCAGACATGGCTGT 350
|||.....|
1320 AlaAsnProLeuGlyGlySerHisGlyAspAlaMetValAsnAlaSer 1336
|||.....|
351 CGTGTACAGGCACTCCACAGACCAGCTGCAGCTCCACAGAGTGGCG 400
|||.....|
1336 uAlaSerLeuAlaSerGlyLeuLysGlnIleProGlnPheAspAspPro 1353
|||.....|
401 TGCAGCAGCAGCAGCAGCAGCAGCTTC..... 429
|||.....|
1353 aLglnGlnSerLeuAlaSerLeuGlnPheSerAlaGlySerThrGlyLys 1369
|||.....|
430 .....CAGCAG.....CAGCAGCAGC 446
|||.....|
1370 SerGlyLeuThrAspAsnPheLeuMetGlnIleHisLeuMetGlnProAl 1386
|||.....|
447 GGGCTACAGCAGCAGCAGCAGCAGCAGCAGCAGTTCAGGCTC 496
|||.....|
1386 aGlyProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 1403
|||.....|
497 AGCAGAGTGCATGCAGCAGCAGCTTCAGACAGTACAGTACAGCAGCAG 546
|||.....|
1403 GlnGlnGln.....GlnGlnGlnGln 1409
|||.....|
547 CAGCTCAGCAGCAGCAGCAGCAGCAGCAGCTA.....ATTAA 587
|||.....|

```











Fri Mar 1 09:11:59 2002

us-09-668-119-1.rsp

Page 16

Ratio: 1.26 Gaps: 35  
Percent Similarity: 48.370 Percent Identity: 30.703

Alignment block:  
US-09-668-119-1 x GLUT4\_WHEAT

Align seg 1/1 to: GLUT4\_WHEAT from: 1 to: 838

```
39  GAATATGGAGAGCAAGTTTCT...GAAGCCAAAGCCGGAGCAAT 85
   |||.....|||
126  G1ylnalaserProclnargProglnglnlnlnlnProglnglnl 142
   |||.....|||
86  ACCTTCTCT...CGTGGC...CAGCTCATATTCATTTCAGACATT 129
   |||.....|||
142  ylnlnlnserlglnglnlnlnlnlnlnlnlnlnlnlnlnlnln 157
   |||.....|||
130  CATACAAAGAAATCTCAAGCTTCGTGATGATCTATGATGACATCA 179
   |||.....|||
158  ....GlnlnlnProglnglnlnlnlnlnlnlnlnlnlnlnlnln 166
   |||.....|||
180  GAG.....CCTGACTGGGAGCCTGCTGGGAGCCGCTGGAATG 220
   |||.....|||
167  Glnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnln 177
   |||.....|||
221  GCATGCTCTCTGGGCGCGGAGAG...TCCTGGGGGG 258
   |||.....|||
178  ....Proglnglnlnlnlnlnlnlnlnlnlnlnlnlnlnlnln 191
   |||.....|||
259  ATGGGTAGCTTGGTGGCATGGGAGAGCAATGCTCTCTCAGGCAACC 308
   |||.....|||
192  GlnProglnglnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnln 202
   |||.....|||
309  GCTCTCTGGAGCCCGAGATGGCCCTCAAGCAGGCTGTGTCTA 358
   |||.....|||
202  oglynglnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnln 213
   |||.....|||
359  CGGCACTCCACAGACCAGCTG...CAGCTCCAGAGTGGCG 399
   |||.....|||
214  ....Serlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnln 225
   |||.....|||
400  CTG...CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 446
   |||.....|||
226  Glnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnln 242
   |||.....|||
447  GGGCTTACAGCAGCAG...CAGCAGCAGCAGCAGCAGCAGCAG 490
   |||.....|||
242  oglynglnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnln 259
   |||.....|||
491  AGGCTAGCAGAGTGCATGACAGCAGCAGTCCAGCAGTGCAGCAG 540
   |||.....|||
259  lnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnln 275
   |||.....|||
541  CAGCAG...CAGCTCCAGCAGCAGCAGCAGCAGCAGCAGCA 575
   |||.....|||
276  Glnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnln 292
   |||.....|||
576  GCATCTAATTAATGATCATCAAAATCAGCAACAGATACAG... 618
   |||.....|||
292  oglynglnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnln 309
   |||.....|||
619  ....CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 633
   |||.....|||
309  yTYrProthSerProglnglnlnlnlnlnlnlnlnlnlnlnln 325
   |||.....|||
634  CTGACAGGATAGCAGCTGACGTCACCAAGCAGCAGCAGCAGCAG 683
   |||.....|||
326  lnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnln 342
   |||.....|||
684  GCAG...CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 727
   |||.....|||
342  nlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnln 359
   |||.....|||
728  CATTCAG...CAGCA...CCGATGAG... 750
```

```
359  roglnglnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnln 375
   |||.....|||
751  CAGCCACAGCCT...CCGCCCTCCAGGCTCTGCC... 783
   |||.....|||
376  Glnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnln 392
   |||.....|||
784  ....CAGCAGCTCCAGCAGATGATCTACACACAGC 814
   |||.....|||
392  lnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnln 409
   |||.....|||
815  ACCACAGCCGACACACAGCCGACAGCCTGCTGACAGCAACCA 864
   |||.....|||
409  lnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnln 423
   |||.....|||
865  CCATCACAACCTCCGACAGCAGCAGCAGCAGCAGCAGCAGCAG 914
   |||.....|||
424  ProglTYrTYrPro...ThSerProleuglnlnlnlnlnlnln 964
   |||.....|||
915  GGAAGCTCTCCGACCAATGTTGATACCAACACCACTGAATTTG 964
   |||.....|||
435  ylnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnln 443
   |||.....|||
965  TCCGAGCTCCGATGTTGTCAGACAGCCGACAGCAGCAGCAGTGC 1014
   |||.....|||
444  ....ThSerProglnglnlnlnlnlnlnlnlnlnlnlnlnln 453
   |||.....|||
1015  CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTGC 1064
   |||.....|||
454  GlnProglnglnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnln 470
   |||.....|||
1065  CCGAGTCCAGTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1114
   |||.....|||
470  ogly...Glnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnln 478
   |||.....|||
1115  CGGCGCAGCAGTGCAGACCCGACGATGATGCCCCGCCCCAGCG 1164
   |||.....|||
478  lnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnln 494
   |||.....|||
1165  TCCCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1214
   |||.....|||
495  Glnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnln 509
   |||.....|||
1215  TGGCCCTGCCATCTCCAGTACCTTCGACCCGACCCGACGAGC 1264
   |||.....|||
509  lnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnln 526
   |||.....|||
1265  CCTCCAGAGCCAGTGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1302
   |||.....|||
526  roglTYrTYrProthSerProleuglnlnlnlnlnlnlnlnln 542
   |||.....|||
1303  TWCAGTCTCCCTCAGCTGACCTTTAAACACAGCAGCAGCAGCAG 1352
   |||.....|||
543  TYrAsPProthSerProglnglnlnlnlnlnlnlnlnlnlnlnln 559
   |||.....|||
1353  TGTGAGAGCCAGCTGCTCCAGCAGCAGCAGCAGCAGCAGCAGC 1395
   |||.....|||
559  nlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnln 576
   |||.....|||
1396  TACCTGACAGCAGTGAAGCAGCTGCAAGTACAGCAGCAGCAGCAG 1445
   |||.....|||
576  lnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnln 591
   |||.....|||
1446  CATGATCAACAAGATCGACAAGAAAGCAAGAAAGAGACCTGAGTA 1495
   |||.....|||
592  Glnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnln 605
   |||.....|||
1496  AGATGAAGAGCCTTGTGACATTCGACAGACCCCTGGAAGCGTGTCC 1545
   |||.....|||
606  ....Glnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnlnln 620
   |||.....|||
1546  CTGAAGACCTTGCAAAAGTGTGAGATGCCCTGAGCAAACTCAAGA 1592
   |||.....|||
```



620 yglIn.....GlnProGlyGlnGlyGln 628

seq\_name: SwissProt\_39:GLT5\_WHEAT

seq\_documentation\_block:

ID GLT5\_WHEAT STANDARD: PRT: 839 AA.

AC P10388;

DT 01-MAR-1989 (Rel. 10, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE GLUTENIN, HIGH MOLECULAR WEIGHT SUBUNIT DX5 PRECURSOR.

GN GLU-1D-1D OR GLU-D1-1B.

OS Triticum aestivum (wheat).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;

OC Triticeae; Triticum.

OX NCBI\_TaxID=4565;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. CHEYENNE;

RX MEDLINE=89098419; PubMed=2563152;

RA Anderson O.D., Greene F.C., Yip R.E., Halford N.G., Shewry P.R.,

RA Malpica-Romero J.M.;

RT "Nucleotide sequences of the two high-molecular-weight glutenin genes

RT from the D-genome of a hexaploid bread wheat, Triticum aestivum L. cv

RT Cheyenne.";

RL Nucleic Acids Res. 17:461-462(1989).

RN [2]

RP REVISIONS, SEQUENCE FROM N.A.

RC STRAIN=CV. CHEYENNE;

RA Anderson O.D.;

RL Submitted (FEB-1997) to the EMBL/Genbank/DBJ databases.

-1- FUNCTION: GLUTENINS ARE THE HIGH MOLECULAR WEIGHT SEED STORAGE

-1- PROTEIN OF WHEAT ENDOSPERM. THOUGHT TO BE RESPONSIBLE FOR THE

-1- VISCO-ELASTIC PROPERTY OF WHEAT DOUGH.

-1- SUBUNIT: DISULFIDE-BRIDGE LINKED AGGREGATES.

-1- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE

-1- GROUP 1 CHROMOSOMES OF WHEAT.

-1- MISCELLANEOUS: THE MATURE PROTEIN IS CHARACTERIZED BY A LARGE

-1- NUMBER OF WELL PRESERVED REPEATS OF THE TWO MOTIFS: GOQPGQ AND

-1- GOQPGQGGQGYPTS.

-1- This SWISS-PROT entry is copyright. It is produced through a collaboration

-1- between the Swiss Institute of Bioinformatics and the EMBL outstation -

-1- the European Bioinformatics Institute. There are no restrictions on its

-1- use by non-profit institutions as long as its content is in no way

-1- modified and this statement is not removed. Usage by and for commercial

-1- entities requires a license agreement (See <http://www.isb-sib.ch/announce/>-1- or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-1- -----

-1- DR EMBL: X12928; CAA31395.1; -

-1- DR PIR: S02262; S02262.

-1- DR InterPro: IPR001419; GLUTENIN.

-1- DR PRINTS: PR00210; GLUTENIN.

-1- KM Seed storage protein; Repeat; Multigene family; Signal.

-1- FT SIGNAL 1 21

-1- FT CHAIN 22 839 GLUTENIN, HIGH MOLECULAR WEIGHT SUBUNIT

-1- FT DOMAIN 131 801

-1- FT SPOUNCE 839 AA; 89359 MM; OF14E110D552643 CRC64;

-1- FT REPEATS.

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1- FT

-1-







```

FT DOMAIN 302 305 POLY-PRO.
FT DOMAIN 376 382 POLY-SER.
FT DOMAIN 386 397 POLY-SER.
FT DOMAIN 442 447 POLY-PRO.
FT DOMAIN 479 483 POLY-HIS.
FT DOMAIN 484 497 POLY-GLN.
FT DOMAIN 504 507 POLY-PRO.
FT DOMAIN 564 574 POLY-SER.
FT DOMAIN 704 707 POLY-PRO.
FT DOMAIN 802 815 ARG/ALA-RICH (MIXED CHARGE).
FT DOMAIN 816 827 ARG/GLU-RICH (MIXED CHARGE).
FT DOMAIN 925 934 ARG/GLU-RICH (MIXED CHARGE).
FT CONFLICT 94 94 MISSING (IN REF. 2).
FT CONFLICT 333 333 Y -> H (IN REF. 2).
FT CONFLICT 339 339 M -> I (IN REF. 2).
FT CONFLICT 541 541 P -> T (IN REF. 3).
FT CONFLICT 1028 1028 G -> A (IN REF. 2).
SQ SEQUENCE 1185 AA; 124785 MW; 56C306267331C005 CRC64;

```

```

alignment_scores:
    quality: 316.50      length: 637
    ratio: 0.995         gaps: 38
Percent Similarity: 49.922 Percent Identity: 27.786

```

alignment\_block:

US-09-668-119-1 x DRPL\_HUMAN ..

Align seg 1/1 to: DRPL\_HUMAN from: 1 to: 1185

```

163 CCGATGAAATGACATCCAGAGCGTACATGGGAGGAGCGGAGCGG 212
    ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
303 ProProProAlaLeuAlaGlyProLeuAlaHisAla...SerAlaSerPro 318
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
213 TGGAAATGGATGCTCTCTCGGGGCCGGGAGACATCTCTGGGGGATGG 262
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
318 oglyleuglyAlaGlyProLeu...ProGlyHis.....LeuPro 330
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
263 GTACGTTGGTGGCATGGGAGACGATGCTCTCTAGGGGAGCGGCT 312
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
330 roSerProGlyAlaMetGlyGlyLeuMet.....GlyGlyLeuPro 343
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
313 CCGTGGG.....ACCTGGGGATGGCCCTCCAGACATGGC 347
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
344 ProGlyProGlyGlyGlyProThLeuAlaProSerProHisSerLeuPro 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
348 TGTGGTGTCTACGGCACTCCAGACACCGAGCTGACCTCCAGAGGTGG 397
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
360 oProAlaSerSerSerAlaPro.....PrometArgPhe 373
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
398 CGGTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGG 447
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
374 ProGlySerSerSerSerSerSerSerAlaAlaAlaSerSerSer 388
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
448 GCGGTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCTCA 497
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
389 .....SerSerSerSerSerSerSerAlaSer.ProProProAlaSer 402
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
498 GCAGAGTGC.....CATGCAGCAGCAGCTCCAGACAGTATGTC 535
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
403 GlnAlaLeuProSerTyProHisSerPheProProProHisSerLeu 419
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
536 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTAT 585
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
419 lValSerAsnGlnProGlyTyThrGlnProSerLeuProSerGlnA 436
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
586 AAATTCATCA..... 596
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
436 lavalITPserGlnGlyProProProProProProTyGlyArgLeu 452
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
597 ...TCAAATTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 643
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
453 AlaAsnSerAsnAlaHisProGlyProPheProProSerThrGlyAlaG 469

```

```

644 TAGCAGAGCTGCAGCTCCA.....ACAGAG 669
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
469 nSerThrAlaHisProProValSerThrHisHisHisHisGln.Gln 485
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
670 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCTTGGAG 719
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
486 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnHisGlyAsn 502
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
720 CCAGCCCAATTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCT 762
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
502 rGlyProPro.....ProProGlyAlaPheProHisProLeuGlnG 516
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
762 ..... 762
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
516 lGlySerSerHisHisAlaHisProGlyAlaMetSerProSerLeuGly 532
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
763 .....CCGCCCTCCAGCT...TGCCCAAGCAGCTGCA 794
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
533 SerLeuArgProTyProProGlyProAlaHisLeuProProProHis 549
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
795 GCAGATGCATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 831
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
549 rGlnValSerTySerGlnAlaGlyProAsnGlyProProValSerSer 566
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
832 CAGCCCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 881
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
566 eSerAsnSerSerSerSerSerSerSerGlnGlySerTy.....ProG 580
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
882 ACAGTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 931
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
581 SerHisProSerProSerGlnGlyProGlnGlyAlaProTyProPhe 597
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
932 AAATGTGTATACCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 981
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
597 oProVal.....ProThrValThrThrSerSerAlaHisLeuSerTy 612
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
982 GTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGT 1031
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
612 allleAlaThrValAlaSerSerPro.....AlaGlyTyThrThr 625
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1032 ACAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1081
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
626 AlaSerProGlyProProProProTyGlyGlyAlaGlyAlaProSer 640
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1082 AGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1121
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
641 .....ProGlyAlaTyThrThrAlaThrProProGlyTyGlyP 654
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1122 .....GCAGTGCAGCAGCAGCAGCAGTGCATGCG 1148
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
654 rGlySerProProSerPheArgThrGlyThrProProGlyTyArgGly 670
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1149 CCCTCCCCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGT 1189
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
671 ThrSerProProAla.GlyProGlyThrPheTySProGlySerProTh 687
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1190 CAGAGCCCAATTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGT 1239
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
687 alGlyProGlyProLeuProProAlaGlyProSerGlyLeuProSer 702
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1240 TTCCTGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1289
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
703 ...LeuProProProProAlaAlaProAlaSerGlyProProLeuSer 718
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1290 GACC.....CCAGAGAACTTCAGTGTGCC...TCAC 1318
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
718 alThrGlnllyGlnGlnlProAlaGlnGlnlTyGlnlThrProGlnSer 735
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1319 CT...GGACCTTTAAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1362
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
735 roValProProAlaValSerProSerProProProGlyValAlaVal 751

```











Fri Mar 1 09:11:59 2002

us-09-668-119-1.rsp

Page 22

```

1456 AAGATGCACAGACGAGACGAGAAAAAGACCTAGTACATGAGAG 1505
378 ..... 378
1506 CCTCTGGACATTTGAGACGACCCCTGCAGCGGTCCCTCAAGACT 1555
379 .....ThrPheThrLeuThrProSerThrAlaThrPro..... 389
1556 TGCAGAAAGTGTGAGATCGCCCTGGAGAAACATCAAGACATGCGGCG 1605
390 .....SerThrAspGlnPheThrGlySer 397
1606 CACATCCCCACAGCCGCCGGTGCACACGCAACACAGCATACT 1651
397 tSerAlaSerThrGlnSerAspSerThrAspSerSerThrValPro 412

```



OM of: US-09-668-119-1 to: SPTREMBL\_17.\* out\_format: pfs

Date: Feb 28, 2002 7:35 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

# Command line parameters:

-MODEL=Irfam+g2p\_model -DEV=xlp  
-O=/cgn2.1/USPTO\_SPOOL/US09668119/runal\_28022002\_084717\_19384/app\_query.fasta.1.1824  
-DB=SPTREMBL\_17 -OPMT=fastan -SUFFIX=rspt -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-GAPOP=6.000 -GAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blowm62  
-TRANS=human40.cdi -LIST=45 -DOCALL=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -MINLEN=0 -MAXLEN=200000000  
-USER=US09668119\_@CGN1\_1.171 -NCPU=6 -ICPU=3 -LONGLOG -NO\_XLPPY  
-WAIT -THREADS=1

## Search information block:

Query: US-09-668-119-1  
Query length: 1740  
Database: SPTREMBL\_17.\*  
Database sequences: 473505  
Database length: 14627329  
Search time (sec): 141.200000

score_list:	Strd Orig	ZScore	Escore	Len	Documentation
Sequence					
SP_human:09P1T3	3010.00	3070.23	1.6e-163	579	! O9P1T3 homo sapiens (human) .
SP_human:015413	1780.00	1820.06	1.2e-93	349	! O54113 homo sapiens (human) .
SP_invertebrate:09Y149	517.00	528.69	4.6e-22	749	! O9Y149 drosophila melanogaster
SP_invertebrate:09YV64	464.50	473.08	4.0e-19	1069	! O9YV64 drosophila melanogaster
SP_invertebrate:09YV67	456.50	460.83	9.8e-19	2123	! O9YV67 dictyostelium discoid
SP_human:014687	447.50	446.60	2.6e-18	4957	! O14687 homo sapiens (human) .
SP_human:014686	447.50	446.25	2.6e-18	5252	! O14686 homo sapiens (human) .
SP_invertebrate:045322	447.00	445.01	4.2e-18	796	! O45322 caenorhabditis elegans
SP_invertebrate:09YV69	444.00	443.03	2.3e-17	900	! O9YV69 drosophila melanogaster
SP_invertebrate:064507	431.00	448.11	4.6e-17	230	! O64507 mus musculus (mouse) .
SP_invertebrate:09YV66	427.50	432.28	4.5e-17	1798	! O9YV66 drosophila melanogaster
SP_invertebrate:062220	427.00	444.22	7.8e-17	223	! O62220 mus musculus (mouse) .
SP_invertebrate:09YV67	423.00	430.65	6.4e-17	1542	! O9YV67 drosophila melanogaster
SP_invertebrate:09YV67	423.00	426.26	7.6e-17	2285	! O9YV67 drosophila melanogaster
SP_invertebrate:09YV61	417.00	425.54	2.1e-16	926	! O9YV61 drosophila melanogaster
SP_invertebrate:09YV62	416.00	421.21	2.1e-16	1612	! O9YV62 drosophila melanogaster
SP_invertebrate:09YV67	410.50	415.98	4.3e-16	1514	! O9YV67 leishmania major .
SP_invertebrate:09YV67	406.00	412.19	8.0e-16	1325	! O9YV67 leishmania major .
SP_invertebrate:09YV67	400.50	408.19	1.7e-15	1013	! O9YV67 drosophila melanogaster
SP_plant:09F41	400.00	412.46	2.3e-15	455	! O9F41 secale cereale (rye) .
SP_invertebrate:013028	399.50	411.30	2.3e-15	507	! O13028 boreogadus salda .
SP_invertebrate:024279	399.00	402.93	1.8e-15	1893	! O24279 drosophila melanogaster
SP_invertebrate:024279	397.00	399.86	2.3e-15	2251	! O24279 caenorhabditis elegans
SP_invertebrate:09YV69	396.50	398.26	2.3e-15	2703	! O9YV69 drosophila melanogaster
SP_invertebrate:09YV64	396.50	394.22	2.0e-15	5322	! O9YV64 drosophila melanogaster
SP_invertebrate:09YV62	394.50	396.16	3.3e-15	1954	! O9YV62 drosophila melanogaster
SP_invertebrate:09YV63	393.50	390.93	2.9e-15	5533	! O9YV63 drosophila melanogaster
SP_invertebrate:024754	388.00	392.53	8.0e-15	1655	! O24754 drosophila virilis (f
SP_human:078364	387.00	394.49	1.0e-14	1004	! P78364 homo sapiens (human) .
SP_invertebrate:09YV67	386.50	383.86	7.3e-15	5476	! O9YV67 drosophila melanogaster
SP_invertebrate:09YV62	386.50	383.80	7.3e-15	5533	! O9YV62 drosophila melanogaster
SP_invertebrate:09YV61	386.50	383.78	7.3e-15	5554	! O9YV61 drosophila melanogaster
SP_invertebrate:09YV61	386.50	383.77	7.3e-15	5560	! O9YV61 drosophila melanogaster
SP_invertebrate:09YV68	385.50	392.91	1.2e-14	1012	! O9YV68 mus musculus (mouse) .
SP_invertebrate:09YV67	383.50	389.09	1.5e-14	1366	! O9YV67 drosophila melanogaster
SP_invertebrate:09YV67	383.50	388.17	1.4e-14	1594	! O9YV67 drosophila melanogaster
SP_invertebrate:09YV69	381.50	381.43	1.4e-14	3502	! O9YV69 drosophila melanogaster
SP_invertebrate:001505	381.00	387.15	2.1e-14	1234	! O01505 canorhabditis elegans
SP_mammal:09G6K7	379.00	390.05	3.4e-14	539	! O9G6K7 equus caballus (horse) .
SP_invertebrate:09YV62	379.00	386.80	3.0e-14	929	! O9YV62 dictyostelium discoid

SP\_invertebrate:09YV63 + 379.00 384.53 2.7e-14 1359 ! O9YV63 drosophila melanogaster  
SP\_invertebrate:09YV63 + 379.00 383.47 2.6e-14 1624 ! O9YV63 drosophila melanogaster  
SP\_invertebrate:046097 + 378.00 382.58 3.0e-14 1589 ! O46097 drosophila melanogaster  
SP\_invertebrate:09YV67 + 378.00 380.35 2.7e-14 2309 ! O9YV67 drosophila melanogaster  
SP\_plant:09YV68 + 374.00 380.36 5.4e-14 1164 ! O9YV68 arabidopsis thaliana

seq\_name: SP\_human:09P1T3

seq\_documentation\_block:

ID O9P1T3 PRELIMINARY; PRT; 579 AA.  
AC O9P1T3;  
DT 01-OCT-2000 (TREMELREL. 15, Created)  
DT 01-OCT-2000 (TREMELREL. 15, Last sequence update)  
DT 01-JUN-2001 (TREMELREL. 17, Last annotation update)  
DE TPA INDUCIBLE PROTEIN.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Abraham S., Solomon W.B.;  
RT "A novel glutamine rich putative transcriptional adaptor protein  
RT preferentially expressed in placenta and bone marrow tissues."  
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF056191; AAC12944.1; .  
DR InterPro: IP0002965; P.Rich-Extensn.  
DR PRINTS: PR01217; PRICHEXTENS.  
SQ SEQUENCE 579 AA; 63879 MW; CCC9E710C6BD02F CRC64;

## alignment\_scores:

Quality: 3010.00 Length: 579  
Ratio: 5.199 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-668-119-1 x O9P1T3 ..  
Align seg 1/1 to: O9P1T3 from: 1 to: 579

1 ATGAGGAAAGCTGCTGTCGACACAGTAATTCAGACAGATTCGAGAG 50  
|||||  
1 MetArgysAlaGlyValAlaHisSerLysSerSerLysAspMetIuse 17  
51 CCATGTTCTCTGAGGCGCAAGCCGCGGAGCAATCTTCCTGCTG 100  
|||||  
17 THLSValPheLeuLysAlaLysThrArgAspGlnYrLeuSerLeuVal 34  
101 CCAGGCTCATTCATTCATTCGAGACATTCATTAACAAGAAATCTCA 150  
|||||  
34 IahArgLeuIleIleHisPheArgAspIleHisLysLysSerGlnAla 50  
151 TCGCTGATGATCTCATTCATTCATTCGACATTCGACATTCGACATTC 200  
|||||  
51 SerValSerAspPrometAsnAlaLeuGlnSerLeuThrIleYrProAl 67  
201 TGGCGGAGCGGCTGAGATTCGATTCGATTCGATTCGATTCGATTC 250  
|||||  
67 AAlaGlyAlaAlaGlyIleIleLysMetProArgGlyProGlyGlnSer 84  
251 TGGCGGAGCGGCTGAGATTCGATTCGATTCGATTCGATTCGATTC 300  
|||||  
84 eugIlyGlyMetGlySerPheGlyAlaMetGlyLysInPrometSerLeuSer 100  
301 GCGGAGCGGCTGAGATTCGATTCGATTCGATTCGATTCGATTCG 350  
|||||  
101 GLyInProProProGlyThrSerGlyMetAlaProHisSerMetAla 117  
351 CGTTCACGACCACTCCACAGACCCAGCTGACCTTCACAGAGTGGCC 400  
|||||  
117 IValSerThrAlaThrProGlnThrGlnLeuGlnLeuGlnValAlaAla 134







```

481 CAGCAGTTCAGGCTCAGCAGAGTCCATGCGCAGCAGAGTTCACAGCAGT 530
    |||||||
    3 GlnGlnPheGlnAlaGlnGln*****AlaMetGlnGlnGlnPheGlnAlaVala 19
531 AGTGCAGCAGCAGCAGCAGCAGCTCCAGCAGCAGCAGCAGCAGCAGCAGC 580
    |||||||
    19 lValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 36
581 TAATTAATTCATCATCAAAATACAGACAGATACAGCAGCAGCAGCAGCAG 630
    |||||||
    36 euileIySleuH1sh1SglnAsnGlnGlnGlnGlnGlnGlnGlnGlnGln 52
631 CAGCTGCAGGGAATAGCAGCAGCTGCAGCTCCACCAACAGCAGCAGCAGCA 680
    |||||||
    53 GlnLeuGlnInrGltleAlaGlnLeuGlnGlnGlnGlnGlnGlnGlnGln 69
681 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGGCTTTGGAGGCCAGCCACCA 730
    |||||||
    69 ngInGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 86
731 TTTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCTCCGCTCCAGGCTCTG 780
    |||||||
    86 lGgInGlnPProMetGlnGlnPProGlnPProPProSerGlnAlaLeu 102
781 CCCAGCAGCAGCTGCAGCAGATGCATCACACACAGCAGCAGCAGCAGCAGC 830
    |||||||
    103 PProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 119
831 ACAGCCCCAGCAGCAGCTCCAGTGTGCTCAAGAACCAACATCAACATCCCGC 880
    |||||||
    119 oGlnPProGlnGlnPProValAlaGlnAsnGlnPProSerGlnLeuPPro 136
881 CACAGTGCAGCAGCAGCAGCTTTGGTGTACAGAGGGCAGAGCTCCCGCTGGA 930
    |||||||
    136 roGlnSerGlnThrGlnPProLeuValSerGlnAlaGlnAlaLeuPProGly 152
931 CAAATGTTGATATACCAACACACACTGAATTTGTCGAGCTCGATGATGT 980
    |||||||
    153 GlnMetLeuTyThrGlnPProLeuValSerGlnAlaGlnAlaLeuPProMet 169
981 GGTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGGTCGACAGCAGCAGCAG 1030
    |||||||
    169 lValGlnGlnPProValAlaGlnPProGlnValGlnGlnGlnGlnGlnGln 186
1031 TACAGCAGCAGCTGCAGCTGCAGATGGTGGCTCCGAGTCCAGCTGAGC 1080
    |||||||
    186 alGlnThrAlaGlnAlaAlaGlnMetValAlaPProGlyValGlnValSer 202
1081 CAGAGCAGCAGCTCCCATGCTGTCTCGCCGCTCACCGGGCCAGCAGGTGCA 1130
    |||||||
    203 GlnSerSerLeuPProMetLeuSerSerSerPProSerPProGlyGlnGln 219
1131 GACCCCGCAGTCGATGCCCCCTCCCGCCAGCCGTCGCCGAGCCGCGGCC 1180
    |||||||
    219 nThPProGlnSerMetPProPProPProGlnPProSerPProGlnPPro 236
1181 AGCCAGCTCAGCAGCAGCAGCTCAGCAGCTGAGCCGCGCCCAATCT 1230
    |||||||
    236 lnPProSerSerGlnPProAsnSerSerValSerSerGlyPProAlaPProSer 252
1231 CCCAGTAGCTCTGCGCCAGCCCTCAACCGAGCCCTCCAGAGCCAGT 1280
    |||||||
    253 PProSerSerPheLeuPProSerPProSerPProGlnPProSerGlnSerPPro 269
1281 GACGGCGGGAGCCCGCAGAGAACTTCAGTGTCCCTCACCTGGACCTTTAA 1330
    |||||||
    269 lThAlaAlaThrPProGlnAsnPheSerValPProSerPProGlyPProLeu 286
1331 ACACACCTGTAGACCCAGCTCTGTGATGAGCCAGCTGAGCTCCAGCAG 1380
    |||||||
    286 snThPProValAlaAsnPProSerSerValMetSerPProAlaGlySerSerGln 302
1381 GCTGAGAGCAGCAGTACTGAGACAGCTGAAGCAGCTGTGGAAGTACAT 1430

```

```

|||||
303 AlAGlUGlUGlNGlNtYrleuAspYrleuYsglnLeuSerYrTyl 319
1431 CGAGCCCCCTGCGCCGCGATGATCAACAGATGCACAGAGAACAGACAGAA 1480
    |||||||
    319 eGlnPProLeuArgArgMetlleAsnIlySleAspYsglnGlnGlnPArgL 336
1481 AAAGAGACCTGAGTAAAGATGAAGAGCCTCTGAGC 1515
    |||||||
    336 yslYAspLeuSerYrlySmetIySserLeuLeuAsp 347
seq_name: sp_invertebrate:09Y149
seq_documentation block:
ID 09Y149 PRELIMINARY; PRT; 749 AA.
AC 09Y149;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE BCDA:GH03922 PROTEIN.
GN ARCI05 OR BCDA:GH03922 OR CG4184.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Planinkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleab J.M.,
RA Palazolo M., Plitman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtkask R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissensbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveli J.S., Zhan M., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RA "The genome sequence of Drosophila melanogaster."
RT Science 287:2185-2195(2000).
RN [2]
RP
RC SEQUENCE FROM N.A.
RA Rubin G.M., Wan K.H., Harvey D., Lewis S.E., Brokstein P., Tsang G.,
RA Agbayani A., Arcina T.T., Baxter E., Blazej R.G., Butenoff C.,
RA Champe M., Chavez C., Chew M., Doyle C.M., Farfan D.E., Frise E.,

```



RA Galle R., George R.J., Harris N.L., Hoskins R.A., Evans-Holm M.,  
 RA Houston K.A., Hummel S.R., Kim E., Li P., Koshnell M., Pacled J.M.,  
 RA Park S., Seguela A., Sethi H., Snit E., Swirskas R.R., Weinburg T.,  
 RA Galkner S.E.,  
 "Full length Drosophila melanogaster cDNA sequence."  
 RL Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF003589; AAF:1490.1; -  
 DR EMBL; AF145620; AAD:8595.1; -  
 DR FlyBase; FBgn002759; Arc105.  
 SO SEQUENCE 749 AA; 80526 MW; 42375AEB/C13A1DF CRC64;

alignment\_scores:  
 Quality: 517.00 Length: 656  
 Ratio: 1.548 Gaps: 26  
 Percent Similarity: 50.915 Percent Identity: 29.268

alignment\_block:  
 US-09-668-119-1 x Q9Y145

Align seg 1/1 to: Q9Y149 from: 1 to: 749

```

25 AGTAATCCAGCAGCATATGAGAGCCATGTTTCTGAAGGCCAAGAC 74
   ::::::::::::::::::::|::|::|::|::|::|::|::|::|::|
30 ThrLysAsnAlaGlyValMetGlnAsnHisLeuPheArgLysSerArg 46
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
75 CCGGAGAGAAATACCTTCTCTGCGGAGGCTCATTTATCCATTTTCAG 124
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
46 TlysAspGlyTyrLeuGlyLeuValAlaLysLeuPheMetHisTyrLys 63
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
125 ACATTTCATAACAAGAAATCTCAAGCTCC..... 153
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
63 spMetSerArgLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 79
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
154 .....GTCAGTGA 161
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
80 GlyProProProAsnAlaGlnMetGlyGlyGlnAsnMetMetGlnAs 96
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
162 TCCTATGATGACATCTCCAGAGCCTGAGCTGCGGAGCGGACGCG 211
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
96 proLeuAsnAlaLeuGlnAsnLeuAla.....SerGlnGlyAsn 110
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
212 CTGGAGATGGCATGCTCTCTGCGGAGCGGAGCATCTCTGGGCGGATG 261
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
110 rGAsnProGlnMetMetProMetGlyAlaGly..... 120
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
262 GGTACTCTTGTGCGCATGGAGAGCCAAATCTCTCTCAGAGGAGCCGCC 311
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
121 .....GlyGlyAlaProVa 125
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
312 TCCTGGGACCTCGGGGATGCCCTCAGACAGATGGCTGTCTGTACGG 361
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
125 lProGlyGlyProGlyThrAlaSerAsn..... 134
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
362 CAATCTCAGACAGCCACTGTCAGCTTCAGAGGTGGCGCTCAGACAGCAG 411
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
135 .....LeuLeuGlnSerLeu 139
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
412 CAGCAGACAGCAGACTTCAGACAGCAGCAGCAGAGCGCGCTACAGACAGA 461
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
140 AsnGlnGlnArgProGlyGlnGln.....MetGlnProMet 152
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
462 GCAGAGCAGCAGCAAGCAGCAGAGTTCAGAGCTTCAGCAGAGTGCATGC 511
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
152 tSerAsnIleArgGlyGlnMetProMetGlyAlaGlyGlyAlaGlyAlaG 169
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
512 AGCAGCAGTTCAGCAGCAGTAGTTCAGCAGCAGCAGCAGCAGTTCAGCAG 561
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
169 lnglnMetMetGln.....ValGlnGlnMetGlnGlnGlyGlyAsnAla 183
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
562 CAGCAGCAGCAGCAGCATCTTAATTAATTCATATCAAAATTCAGACAGA 611
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
184 ProGlyValMetAsnValMetGlyAlaGlyGlyGlnAsnGlnGlnGly 200
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

612 GATACAGCAGCAGCAGCAGCAGCTGTCAGGAAATAGCAGAGCTGCAGCTCC 661
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
200 nIleValGlyAsnProGlyGlnGlnMetGlyValGly..ValGlyMetPr 216
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
662 AACCAACAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCT 711
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
216 oAsnGlnMetValGlyProGlyProAsnSerGlyProAlaValAlaGlyVa 233
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
712 TTGGAGGCGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 761
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
233 lAglyGlyPro.....AsnAlaAla..ProGlyAlaGlyGlyProGlyPr 247
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
762 TCCGCGCTCCAGAGCTGCGCCAGCAGCAGCTGCGAGAGATGATCAGACAC 811
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
247 oAsnGlnMetGlnGlyGlyProMetAsnValAlaAlaMetGln..... 261
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
812 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCTGCTCAGAAC 861
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
262 .....GlnMetProProMetGlnGln 269
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
862 CAACCATCAGCACTCCCGCAGAGTCGACAGCCAGCAGCTTGTGTACAGA 911
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
270 GlnGlnAsnGlnLeuGlyMetGlyMetAsnProMetMetArgMetGlyG 286
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
912 GCGC.....CAAGCTCTCCCTGGCAAAATGTTGTATA 943
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
286 nGlyAsnGlyMetGlyGlyProGlnGlnGlnMetProGlyGlnGlnGlnG 303
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
944 CCCAGCCA.....CCAGTGAATTGTCCGAGCTCGAGATGCTGTGTCAG 987
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
303 lMetProGlnGlnProHisAsnValValGlyGlyProAlaGlyGlnGln 319
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
988 CAG.....GTCAG..... 1002
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
320 GlnValGlyGlyAlaGlyLeuProProAsnAlaValGlnGlnGlyGlyMe 336
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1003 .....CCCGAGGTGCGAGCAGCAGC 1021
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
336 tAsnProMetGlyGlyMetGlyValAsnMetProProAsnLeuGlnGln 353
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1022 AGACAGCAGTACAGACAGCTCAGGCTGCCAGATGCTG..... 1059
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
353 ySerProAsnMetProMetGlyGlnAlaGlyGlnMetPheProGlyAsnArg 369
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1060 ...GCTCCGAGGTCCAGGTCCAGCAGCAGAGCAGCTCC...ATGCTGTC 1103
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
370 GlyGlyValGlyValGlyGlyGlnGlnProGlyGlnProPheMetArgSe 386
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1104 CTCGCGGTCCAGCGGC..... 1119
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
386 rSerProSerProAlaAspAlaGlnGlnGlnGlnGlnAlaGlnLeuc 403
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1120 .....CAGCAGGTG.....CAGACCCCG 1137
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
403 lnglnMetGlnGlnGlnGlnGlnGlnValValGlyAsnGlnThrPro 419
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1138 CAGTCAGATGCCCCCTCCGCCAGCGCTCCGCCAGCCGCGGAGCGCCAG 1187
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
420 ThrGlnGlnProProThrProGln..... 427
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1188 CTCAGAGCCCACTCCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 1237
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
428 ...MetProThrProAsnMetIleProSerProAlaLeuValProGlnS 443
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1238 GCTTCTCGCCAGCCCTCAGCAGCAGCTCCAGAGGCCAGTACAGCGCG 1287
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
443 rSerProGlnMetMetGlnMetGlnAsnSerGlnArgAsnIleArgGln 459
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1288 CGGACCCCAAGAACTTCAAGTGTCCCTCAGCTGAGCTTGA.....AA 1331
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
460 GlnSerPro.....SerAlaSerIleAsnThrProGlyGlnValThrGlyAs 475
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```















DR PROSITE; PS50280; SET; 2.  
 DR SMART; SM00542; EYR; 1.  
 DR SMART; SM00541; EYR; 1.  
 DR SMART; SM00398; IMG; 1.  
 DR SMART; SM00249; PHD; 4.  
 DR SMART; SM00508; POSTSET; 1.  
 DR SMART; SM00317; SET; 1.  
 SO SEQUENCE 4957 AA; 531848 MW; 1026562E1419CEBD CRC64;

alignment\_scores:  
 Quality: 447.50 Length: 647  
 Ratio: 1.387 Gaps: 30  
 Percent Similarity: 43.586 Percent Identity: 29.675

alignment\_block:  
 US-09-668-119-1 x 014687 ..

Align seg 1/1 to: 014637 from: 1 to: 4957

```

226 CCTCTCGGGGCGGCGAGAGTCTGTGGCGGAGTGGTGGTGC 275
|||||:|||||
3214 ProProGInGlyProGInGlyMetLeuGly..... 3223
276 CATGGAGACGCAATGTCTCTCAGGCGCGCGCTCTGGACCTCG 325
:::|||||
3224 .ProAlaGInValAlaValLeuGInGInGInHisProGlyAlaLeuGlyP 3240
|||||:|||||
326 GGATGGCGCTCAGACATGGCTGTGGTGTAGCGACCTCCAGACCC 375
|||||:|||||
3240 roGInGlyProHisArgInValLeuMetThrGInSerArgValLeuSer 3256
:::|||||
376 CAGCTGAGCTTCAGACAG.....GT 395
:::|||||
3257 SerProGInLeuAlaGInGInGInGlyLeuMetGlyHisArgLeuVal 3273
|||||:|||||
396 GCGCGTGCAGACGACGACGACGACGAGTTCAGACGACGACGAG 445
:::|||||
3273 LThrAlaGInGInGInGInGInGInGInGInGInGInGInGInGInGIn 3290
446 GCGCG..... 450
3990 eTAlaGlyLeuSerHisLeuGInGInSerLeuMetSerHisSerGlyIn 3306
451 .....CTACAGCAGCA 461
3307 ProGlySerAlaGInProMetGlySerLeuGInGInLeuGInGInGIn 3323
|||||:|||||
462 GCAGCAGCAGCAGCAGCAGCAGCAGTTCAGACGAGTTCAGATGC 511
|||||:|||||
3323 nGInGInLeuGInGInGInGInGInGInGInGInGInGInGInGInGInGIn 3337
|||||:|||||
512 AGCAGCGCTCCAAAGCAGTGTAGCAG.....CAGCAGCAGCAGCTCCAGCAG 558
|||||:|||||
3337 InGInGInLeuGInGInGInGInGInGInGInGInGInGInGInGInGIn 3353
|||||:|||||
559 CAGCAGCAGCAGCAGCAGCAGCAGTTCAGACGAGTTCAGACGAGCA 608
|||||:|||||
3354 GInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGInGIn 3366
|||||:|||||
609 ACAGATACAGCAGCAGCAGCAGCAGTTCAGACGAGTTCAGACGAGC 658
|||||:|||||
3366 nGInLeuGInGInGInGInGInGInGInGInGInGInGInGInGInGInG 3383
659 TTCACACAGCAGCAACAACAGCAGCAGCAGCAG..... 690
|||||:|||||
3383 InGInGInProGInGInGInGInGInGInGInGInGInGInGInGInGInGIn 3399
691 .....CAGCAGCAGCAGCAGCAGCAGCTTGGCA 716
3400 GInSerArgThrLeuSerProGInGInGInGInGInGInGInGInGInGIn 3416
717 GCGCCAGCCAGCAATTCAGCAGCAGCAGCAG..... 750

```

```

3416 AlaGlyProGInMetProAlaLysProLeuGInHisPheSerSerProG 3433
750 ..... 750
3433 LValAlaGlyProThrLeuLeuLeuThrGlyGlyGInAsnThrVal 3449
750 ..... 750
3450 AspProAlaValSerSerGluAlaThrGInGlyProSerThrHisGIn 3466
750 ..... 750
3466 YGlyProLeuAlaLeuGlyThrThrProGInSerMetAlaThrGInProG 3483
750 ..... 750
3483 LysIValLysProSerLeuSerGlyAspSerGInLeuLeuValGIn 3499
751 ...CAGCCACAGCTCCGCCCTCCAG.....GC 776
3500 ProGInProGInProGInProSerSerLeuGInLeuGInProProLeuArg 3516
777 TGTGCCCGCAGCAGCTCAGCAG.....ATGCAT..... 804
3516 GLeuProGInGInGInGInGInGInGInValSerLeuHisThrAlaGlyG 3533
804 ..... 804
3533 LysGlySerHisGInLeuGInGlySerGlySerSerSerGluAlaSerSer 3549
804 ..... 804
3550 ValProHisLeuLeuAlaGInProSerValSerLeuGlyAspGInProG 3566
805 ...CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 850
3566 YSerMetThrGInAsnLeuLeuGlyProGInGInProMetLeuGInValArg 3583
851 TGTCTCAGAACCAACCATCAGCAACTCCGCGCAGCTCCAGACCCAGCTT 900
3583 rMetGInAsnAsnThrGlyProGInProPro.....LysProGlyPro 3597
901 TTGGTGTCCAGAGCCGACAGCTCTCCCTGGACAAATGTTGTATACCAAC 950
3598 ValLeuGInSerGlyGInGInGInGInGInGInGInGInGInGInGInG 3612
951 ACCACTGAATTTGTCCAGCTCCGATG.....GTGGTCAGCAGCCGCC 994
3612 oThrValGlyGInLeuArgAlaGInLeuGInGInGInGInGInGInGInGIn 3629
995 CA.....GTCCAGCCCGCAGTCCAGCAGCAG..... 1020
3629 roGInLeuArgHisLeuSerProGInGInGInGInGInGInGInGInGInGIn 3645
1021 .....CAGACAGCAGTACAGACAGCTCAGGCTGCCAGATGGG..... 1059
3646 LeuMetGInArgGInLeuGInGInSerGInAlaValArgGInThrProPr 3662
1060 .....GTCGCCGAGTCCAGTCCAGCAG.....CAGACAGCAGCTC... 1092
3662 oTyGInGInLysProGInThrGInGInGInSerProLeuGInGInGInGInG 3679
1093 .....CCCATGCTGTCTCCGCGTCCAGCGGCGCAGCAGAGTTCAGACCCG 1137
3679 YSGInProGInLeuGInGly...PheProGlyProGInThrGInProLeu 3694
1138 CAGTGCATGCCCTCCCGCCAGCAGCTCCCGCAGCAGCCGCGCAGCCAG 1187
3695 GInGInLeuGInGInGInGInGInGInGInGInGInGInGInGInGInG 3711
1188 CTCACAGCCCAATCTCAACAGTCCAGCTCCGCG..... 1218

```







```

3738 1yAlaLeuGlyProThrLeuLeuLeuThrGlyGlyGluGlnAsnThrVal 3754
750 .....
3755 AspProAlaValSerSerGluAlaThrGlnGlyProSerThrHisGlnI 3771
750 .....
3771 yGlyProLeuAlaIleGlyThrThrProGluSerMetAlaThrGlnProG 3788
750 .....
3788 1yGluValLysProSerLeuSerGlyAspSerGlnLeuLeuValGln 3804
751 ...CAGCCAGACCTCCGCCCTCCAG.....GC 776
3805 ProGlnProGlnProGlnProGlnProGlnProGlnProGlnProGln 3821
777 TCTGCCCCAGACGCTCCAGCAG.....ATGCAT..... 804
3821 GlnProGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 3838
804 .....
3838 1yGlySerHisGlyGlnLeuGlySerGlySerSerSerGluAlaSerSer 3854
804 .....
3854 ValProHisLeuAlaGlnProSerValSerLeuGlyAspGlnProGln 3871
805 ...CAGACAGACCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 850
3871 ySerMetThrGlnAsnLeuGlyGlyProGlnProMetLeuGlnArgp 3888
851 TTGGCTCAGACCAACATCACTCCGCGCAGCCAGCCAGCCAGCCAGCC 900
3888 rometGlnAsnAsnThrGlyProGlnProPro.....LysProGlyPro 3902
901 TTGGTGCAGACGGCGAAGCTCCCTGAGCAATGTTGATATCCCAAC 950
3903 ValLeuGlnSerGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 3917
951 ACCACTGAATTTGTCGAGCTCCGATG.....GTGGTCAGACCAACCC 994
3917 oThrValGlyGlnLeuArgAlaGlnLeuGlnGlnGlnGlnGlnGlnGln 3934
995 CA.....GT3CAGCCCGAGGTGAGCAGCAG..... 1020
3934 roGlnLeuArgHisLeuSerProGlnGlnGlnGlnGlnGlnGlnGlnGln 3950
1021 ...CAGACAGCAGTACAGACAGCTCAGCAGTCCAGATGGTG..... 1059
3951 LeuMetGlnArgGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 3967
1060 .....GCTCCCGGATCCAGGTCCAGC.....CAGACAGACCTC... 1092
3967 oTyTGlnGlnProGlyThrGlnThrSerProLeuGlnGlnGlnGlnGln 3984
1093 ...CCAGTGTGTCTTCGCGCTCCAGCCAGCCAGCCAGCCAGCCAGCC 1137
3984 ysglnProGlnLeuGlyGly...PhaProGlyProGlnThrGlyProLeu 3999
1138 CAGTCAGTCCAGCTCCAGCTCCAGCTCCAGCTCCAGCTCCAGCTCCAG 1187
4000 GlnGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 4016
1188 CTCACAGCCCACTCCAGCTCCAGCTCCAGCTCCAGCTCCAGCTCCAG 1218
4016 a.....ProProGlyAlaLeuSerThrGlyProValLeuGlyProVal 4031
1219 ..CCTGCCCATCTCTAGTCTCTGAGCTCCAGCC...CTCAGCCGAG 1263

```

```

4031 IsProThrProProProSerSer..ProGlnGlnProLysArgProSerG 4047
1264 CCTCCAGAGCCCACTGACGCGCGAGCCAGCCAGCACTGCTGCTCC 1313
4047 InleuProSerProSerSerGlnLeu...ProThrGlnAlaGlnLeuPro 4062
1314 CTCACCTGACCTTTAAGACACACCTGTAACCCAGCTGTCATGAGCC 1363
4063 Pro.....ThrHisProGlyThrProLysProGln.....GlyPr 4074
1364 CAGCTGGCTCCAGCCAGCGCTGAGGACGACGACGACGACGACGACG 1413
4074 oThr...LeuGlnProProGlyArgValSerProAlaAlaAlaGlnL 4090
1414 CAGCTGTCAGAT...CATCGAGCCCTGCGCGCGCATGATCAGACGAT 1460
4090 euAlaAspThrLeuPheSerLysGlyLeuGlyProThrPaspProPasp 4106
1461 CGACAAAGAAACAGACAGAAAGAGCCCTGAGTATGATGAGACCTTC 1510
4107 AsnLeuAlaGlnThrGlnLys.....ProGlnGlnSerSerLeuValPr 4121
1511 TGACATCTCTGACAGACCCCTGCAAGCGGTGTCCTGAA 1550
4121 oGlyHisLeuAspGlnValAsnGlyGlnValValProGln 4134

```

```

seq_name: sp_invertebrate:045322
seq_documentation_block:
ID 045322 PRELIMINARY; PRT; 796 AA.
AC 045322:
DT 01-JUN-1998 (Tremblere, 06, Created)
DT 01-JUN-1998 (Tremblere, 06, Last sequence update)
DE 01-JAN-1999 (Tremblere, 09, Last annotation update)
DE DYS 5.5. PROTEIN.
GN DY5.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Lennard N.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RX [2]
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kersey J., Kirschen J., Kirschen J., Lettice P.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Smalton N., Smith A., Sonnenhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan R., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Spoat J., Wolfman P.,
RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans".
RI Nature 368:32-38(1994).
DR EMBL, Z96047; CAB09418.1; -.
SQ SEQUENCE 796 AA; 86783 MW; 4FDFCB8B5461059 CRC64;

```

```

alignment_scores:
Quality: 447.00 Length: 509
Ratio: 1.781 Gaps: 24
Percent Similarity: 49.312 Percent Identity: 32.024

```

```

alignment_block:
US-09-668-119-1 x 045322
Align seg 1/1 to: 045322 from: 1 to: 796

```



```

967 GAGAGCTGG.....ATGGCGT 983
968 :::: |||
969 Gln1LeuProthHisSerThySerTyrlGlyAsnIstGlySerValII 565
970 :::: |||
984 GCACAGACCCCGGAGTCAGACCCCGAGTCAGACAGCAGACAGCAGTAC 1033
985 :::: ||| :::: |||
986 egluGluGluProHisIstGlySerValIIleSerGlnGln.....598
1034 AGACAGCTCAGGCTGCCAGATGGTGGCTCCCGGAGTCSAGTACGCCAG 1083
1035 :::: |||
599 .....ValIaIPro..... 601
1084 AGACGCTCCCGCATGCTCTCTCGCGCGTCACGGGCGCAGAGTGGCAGC 1133
1085 :::: ||| :::: ||| ||| :::: ||| ||
602 .....ProArGProValGlnProValProProMetSerValaIaTh 615
1134 CCGCAGTCAGTACCCCTCTCCCGCCAGCCGTCGCCGAGCGCGCAGC 1183
1135 :::: ||| ||| ||| ||| |||
615 rValThArGglYProProProProThr.....Arg 626
1184 CACAGTCA.....CAGCCCAACTCCAAAGTCAGCTGTGGCCCTGCCCA 1227
1185 :::: ||| ||| :::: ||| |||
626 roAlaSerValIIleAlaProSerSerSerValIaI.....ProLysSer 640
1228 TCTCCCACTAGCTTCTCGCCCAAGCCCTTCACCGCAGCCCTCCAGAGCC 1277
1229 :::: ||| ||| ||| ||| :::: |||
641 GlnProArG.....ProSerProThrThrArGProProThrArGAl 654
1278 AGTACAGCGCGGACCCCAAGACCTTCAAGTGGTCCCTCA...CTGGAC 1324
1279 :::: ||| ||| ||| ||| ||| :::: ||| ||| ||| |||
654 aThIIleuYProThrProGlnAsnIstGlyAlaValThSerArGProAla 671
1325 CTTTAAACACACCTGTGAAACCCAGCTCTGTCAAGAGCCACAGCTGGCTC 1374
1326 :::: ||| :::: ||| ||| ||| ||| :::: ||| |||
671 roAlaThr.....SerSerValIaIThrProSerProSer 682
1375 AGCAGAGCTGAGAGAGAG.....CAGTACCTCGA 1403
1376 GlnSerLeuThrGlnIstGlnIleArGlyLeuProAlaValLeuTyIleAs 1446
1404 CAACCTGACAGCAGCTGTGCAGATACATGAGCCCTCGGCGCG..... 1446
1405 :::: ||| :::: ||| ||| ||| |||
699 pSerLysAsnGlnSerThrLysLysThsGlnThrLeuLeuArGAspThrT 716
1447 .....ATGATCAACAGATGACAGACAGAA 1473
716 YrGlyLeuProLeuValIThrPheTyValaIAspLysThrAspLysProAla 732
1474 GACAGAAAAAGACCTGAGTAAAGATG 1500
1475 :::: ||| ||| ||| |||
733 AlalIeGlnArGlnIleuGlnGlnIleu 741

seq_name: sp_invertebrate:Q9VKK09
seq_documentation_block:
ID Q9VKK09 PRELIMINARY; PRT; 900 AA.
AC Q9VKK09
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CG6043 PROTEIN.
GN CG6043.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY.
RA MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananthites P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

```















```

336 GlnProGlnGlnGlnGlnGlnProGlnGlnGlnHisThrProse 352
333 CCCACAGAGATGCGTGTCTACGACACTCCACAGACC..... 375
352 rProArgGlnSerProLeuGlnProThrThrProThrLeuGlnG 369
376 .....CAGCTGCAGCTCCAGCAGGTGGCG 399
369 GlnGlnProAsnGlnGlnAsnAlaGlnGlnGlnGlnGlnGln 385
400 CTGACAGACAGACAGACAGACAGACTTCCAGACAGACAGCGGC 449
386 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 402
450 GCTACAGACAGACAG.....CAGCAGACGAA..... 477
402 lLeuThrGlnGlnProGlnProGlnGlnGlnGlnValIleThrG 419
478 .....CAGCAGACTTCAG 492
419 lnaArgHisValIleAsnThrSerThrAlaGlnGlnGlnIleGln 435
493 GCTACAGACAGTGCATGACAGACAGACTTCAAGACAGTGCAGAGA 542
436 SerHisMetSerLeuAlaLeuGlnGlnGlnGlnGlnGlnGln 452
543 GCAGACACTCCAGACAGACAGACAGACTCTAATTAAATTGC 552
452 nGlnGlnAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 469
593 ATCAT.....CAAAATCAGACAGATACAGACAGACAGACAG 630
469 lGlnLeuProProAlaGlnGlnGlnGlnGlnGlnGlnGlnGln 485
631 ...CAGCTGCAGCAATAGCAGCTGCAGCTCAACACAGCAA..... 672
486 ValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 502
672 ..... 672
502 eAlaLeuGlnAlaGlnGlnGlnValArgIleGlnGlnThrIleGln 519
673 .....CAGCAGACAGACAGACAGACAGACAGACAGACAG 705
519 rProProThrAsnValGlnGlnGlnGlnGlnGlnGlnGlnGln 535
706 CAGGCTTTGGAGCCAGCCAGCCAAATTCAGACAGCCAGTGCAGAGCC 755
536 GlnGlnIleIleGlnGlnProGlnGlnGlnGlnGlnGlnGln 552
756 ACAGCTCCGCGCTCCAGGCT...CTGCCAGCAGTGCAGCAATGC 802
552 oGlnGlnGlnIleSerGlnValGlnGlnGlnGlnGlnGlnGln 569
803 ATCACA.....CAGCAG...CAGCAGCCGCGCAGCAG 834
569 lnyThrPheIleIleSerGlnGlnHisPheAsnGlnLeuSerValGln 585
835 CCCAGCAGCAGCTCCAGTGTCTAGAAACAGCAGTCAAA..... 873
586 GlnGlnGlnGlnIleLeuAlaGlnAsnProGlnSerGlnAsnValIlyIva 602
873 ..... 873
602 lValAsnGlnThrGlnGlnGlnLeuValGlnGlnGlnGlnGlnGln 619
874 .....CTCCGCGCAGTGCAGACAGCCAGCTTTGGTGTGCAG 912
619 lnyGlnIleIleProGlnGlnGlnGlnGlnGlnGlnGlnGln 633
913 GCGCAGACTCTCCCTGCAGCAAAATGTGTACGCAA..... 948

```

```

634 MetGlnGlnGlnProGlnGlnIleValValAsnGlnGlnIleLeuSerAs 650
949 ...CCACACTGAATTTGTCCGAGCTCCGATGCTGTCAG...CAGC 991
650 pProGlnArgLeuGlnIlyLeuGlnGlnGlnGlnGlnGlnGlnGln 667
992 CCCAGTGCAGCGCCAGTGCAGCAGCAGCAGCAGCAGTACAGACACT 1041
667 lnyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 683
1042 CAGGCTGCCAGATGGTGGCTCCCGAGTCCAG.....GTCAAGCA 1082
684 GlnGlnGlnGlnLeuValGlnGlnGlnGlnGlnGlnGlnIleLeuIleGln 700
1083 GAGCAGCTCCCATG.....CTGCTCTGCG 1108
700 nGlnGlnThrProGlnGlnGlnGlnIleLeuGlnGlnGlnGlnSerAsnA 717
1109 CGTCACCGCGCCAGCAGTGC.....CAGACCCCG 1137
717 sPAsnGlnGlnGlnGlnIleMetGlnLeuIleThrGlnGlnGlnLeuPro 733
1138 CAGTGCATGCCCTCCCTCC..... 1155
734 GlnArgGlnProProGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 750
1156 .....CCGACGCTCCCGCAGCGCGCCAG..... 1182
750 nGlnGlnGlnGlnSerProGlnHisHisProGlnMetGlnGlnGlnHisT 767
1183 ....CCGACGTCAGCCCAACTCAGGTCAGTGGCTGGCTGGCCCA 1227
767 rPserProGlnSerProAlaGlnGlnGlnGlnGlnGlnGlnGlnGln 783
1228 TCTCCAGTAGCTTC.....CTGCC.....AG 1250
784 ThrProThrSerValGlnMetGlnSerProLeuProGlnGlnGlnGln 800
1251 CCCCTCAGCCGAGCCCTCCAGAGCCAGTGAAGGCGGAGCCGACAGA 1300
800 rProGlnProGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 817
1301 ACTTCAGTGTCC.....TCAGCTGAGCTTTAAACAGACT 1338
817 lnyThrSerThrProGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 833
1339 GTGAACCCAGCTCTCATAGCCAGCTGCTCCAGCCAG...GTGA 1385
834 GlnGlnGlnGlnIleValLeuProProProGlnGlnGlnGlnGln 850
1386 GGAGCAGCAGTACCTGGAACAAGCTGAAGCAGCTGTGAAATGACAGC 1435
850 nGlnGlnGlnHisLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 866
1436 CCCTGCGCGCATGATCAACAAGATGACAGAGAAAGCAAGAAAGAAAG 1485
866 roGlnGlnGlnIlyIleValValAspIlyIlyThrHisAlaHisLeuMet 882
1486 GACCTGAATGAATGAAGAGCTTCTGAGCATCTGACAGACCCCTGAA 1535
883 SerLeuAspAlaHisIlyArg...AlaGlnIlyThrIleThrIlyLeuAsnGln 898
1536 GCGGTGTCCCTGAAGACCTTGCAGAAAGTGTGAATGCGCTGTGAAGAAC 1585
898 nGlnAsnIlyProArgValMetLeuArgGlnGlnGlnGlnGlnGlnGln 913
1586 TCAAGAAATGACATGCGGTGCCACTCCGACAGCCGCGGTGCAGAGAA 1635
914 .....Pro.ArgProGlnAlaThrProGlnGlnGlnGlnGln 923
1636 CCAACAGCAGTACCTATGCGAGCGGCTCTGATGCGCTGTGGGCAAC 1685
923 lAlaThrArgProAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 940

```















```

961 OHISGLuIn.....GlnProGlnGlnGlnGln 972
386 TCCAGCAGGTGGCGTCGACAG..... 408
972 roLeuGlnGlnGlnLeuGlnLeuMetHisThrAsnValGlnAlaPro 988
409 .....CAGCAGCAGCAGCAGCAGTTCACAGCAGCAGCAGCGGCGCT 452
989 AspLeuThrGlnGlnGlnMetAlaGlnGlnAlaGlnGlnLeuThr 1005
453 ACAGCAGCAGCAGCAGCAGCAGAA.....CAGCAGCAGT 487
1005 eGlnGlnGlnGlnGlnGlnProGlnAlaValAlaAsnMetGlnGlnAla 1022
488 TCCAGGCTCAGCAGAGTGCATGACAGAGCTTCACAGCAGTACAG 537
1022 yrrAlaMetGlnGlnAlaGlyGlnGlnGlnGlnLeuSerGlnProLeuGln 1038
538 CAGCAGCAGCAGCTCCAGCAGCAGCAGCAGCAGCAGCAGCTAAATTAA 587
1039 lLeGlnGlnGlnLeuGlnGlnGln.....ValAl 1050
588 ATTGCATCATCAAAATCAGCAGCAGTACAGCAGCAG.....CAAGCAGC 631
1050 aValSerHisGln...GlnGlnlLeMetGlnGlnGlnLeuAlaGlnHisG 1066
632 AGCTGCAGCAATATGACAGCTGCATCCAAACAGCAGCAGCAGCAG 681
1066 lInLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1082
682 CAGCAGCAGCAGCAGCAGCAGCAGCAGGCTTGGAGCCACAGCAGCAAT 731
1083 GlnGlnGlnlLeGlnGlnGlnGln.....Le 1091
732 TCAGCAGCAGCAGTGCAGCAGCAGCAG.....CCTCCGCCCTCCAGG 775
1091 uGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1108
776 CTCTGCCCCAGCAGCTGCAGCAG.....ATG 801
1108 lAeMetProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1124
802 CATCAGCAGCAGCAGCAGCAGCG..... 825
1125 AlAProHisGlnHisGlnGlnGlnProGlnGlnGlnGlnGlnGlnGln 1141
826 .CCAGCAGCAGCAGCAGCAGCTCCAGT..... 852
1141 lProProThrSerValAlaProProGlnGlnHisThrTyraGlnGlnGln 1158
853 .....GCTCAGAACCA..... 864
1158 lYglYglValThrLeuSerAspAlaGlnGlnGlnGlnHisProGlyPhe 1174
865 .....CCATCACAACCTCCGCCA.....CAGTCGACAGCAGCAG..... 897
1175 SerAlaValProGlnGlnAlaAlaProPheGlnGlnGlnProThrGlnGln 1191
898 .....CCTTGGTGTCACAGCGCAGCAGCTCTCCCTG 928
1191 nProGlnGlnLeuSerMetProLeuGlnGlnGlnGln..... 1204
929 GACAAATGTTGTATACCAACCA..... 951
1205 ..GlnLeuLeuHisSerGlnProAlaGlnGlnGlnGlnAlaMetSerGln 1220
952 .....CCACTG...AAATTGTCGAGCTCCGATGGTGTGCA 986
1221 GlnGlnGlnGlnGlnProLeuValGlnGlnGlnGlnGlnGlnGlnGln 1237
987 GCAGCCCCCAGTGCAGCCCCCAGTGCAGCAGCAGCAGCAGCAGTACGA 1036
1237 nGlnProProLeu.....ValGlnGlnGlnGlnGlnGlnGlnGlnGln 1251

```

```

1037 CAGCTCAGGCTCCCAAGATGTCGCTCCCGAGATCCAGGTACGCCAGC 1086
1251 lAsGlnGln.....ProSerValGlnHisGlnGln... 1260
1087 AGCCTCCCATGCTGTGCTCCCGGTCACCGGCGCAGGTGCAGAACCC 1156
1261 .....ProLeuValGlnGlnProGln.....GlnGlnGlnProGlnPr 1273
1137 GCAGTCGATGCCCTCCCTCCCTCC.....CAGC 1162
1273 oGlnAsnGlnGlnProGlnProGlnGlnGlnGlnGlnGlnGlnGlnGln 1290
1163 CGTCCCGCCAGCGCCGCGCAGCCCACTACAGCCCACTCCAGCTCAGC 1212
1290 lInGlnProProGlnGlnGlnGlnProAlaValGlnGlnGlnGlnGln 1306
1213 TCTGCGCCTGCGCCATCTCCAGTGTTCGTCGCGCCAGCCCTCAGCGCA 1262
1307 Ser.....GlnValProValGlnGln 1313
1263 GCCCTCCAGAGCCCAATGACGGCGCGAGCCCAACAGAACTTCAGTCTCC 1312
1313 nGlnAsnLeuGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1330
1313 CCTACCTGGAACCTTTAAACACA.....CCTGTGAACCCCAAGCTCT 1353
1330 sPAlaMetSer...LeuAsnSerAlaHisGlyAlaLeuGlnProAlaPro 1345
1354 GTCATAGAGCCCGAGCTGCAGCAGCAGCTGAGGACAGCAG..... 1395
1346 lYsThrGlnProGlnGlnAsnSerAlaAspAlaGlnGlnGlnGlnGln 1362
1395 ..... 1395
1362 rGlyThrGlyThrArgSerGlnGlnGlnProArgArgSerAsnArgSerGly 1379
1396 .....TACCTGGCAACACTCAACAGCAG 1416
1379 snGlnArgThrProLysLeuSerValThrSerValAspGlnGlnGlnGln 1395
1417 CTGTGGAAGTACATCAG.....CCCTGCGCGCGCATGATCAACAGATGCA 1463
1396 lLeAsnGlnHisMetGlnAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1412
1464 C.....AAGACGAGACAGAAAAAAGACCTGAGTACATG.....A 1501
1412 rTLeGlyAspArgAspGlnThrAlaSerAspTLeThrLysMetPheGln 1429
1502 AGAGCCTTCTGACATCTGTACAGACCCCTGGAAGCGGTGTCCCTGAG 1551
1429 roTThrPheHisGlyValGlnThrLeuProSerGlyGlyGlyAlaGln 1445
1552 ACCTTGCAAAAGTGTGAGATCGCCCTGGAGAAACTCAAGATGACATGCG 1601
1446 Gln...SerAsnGlnAsnLeuThrLeuGln.....AlaArg 1456
1602 GTGCGCCACTCCCGCCAGCGCCCGCGGTGCACCGACCA.....AAGACA 1645
1456 gSerHisLeuSerAsnLeuProAsnAlaLysGlnProGlnLeuAsnVal 1473
1646 GTACCTATGCGCAGCGCGCTCGATGCGCGTCTGGCCCAACATCGGCTCAG 1695
1473 eThr.....ProProThrThrThrSerThrMet 1482
1696 CTGTCTTCACAC 1707
1483 SerSerSerSer 1486

```

```

seq_name: sp_invertebrate:O9W3G1
seq_documentation_block:
ID O9W3G1 PRELIMINARY; PRT; 926 AA.

```











Fri Mar 1 09:12:00 2002

us-09-668-119-1.rpt

---

Page 22







215 CCATGTTTCTCGAAGGCCAGACCCGGGACGAAATACCTTCTCTCGTGG 264  
 34 IAAATGIEUILLIEHH:spheargaspiliehiasenylserserinala 50  
 265 CCAGGCGCATATATCCATTTTGGAGACATTCATACAGAAATATGCAAGCT 314  
 51 SerValSerAspProIetAsnAlaLeuGlnSerLeuThrGlyProAl 67  
 315 TCCGCGATGATCCTATGAAATGCACTCCAGAGCTGACTGGCGGACCTGC 364  
 67 aAlaGlyAlaAlaGlyTlIeGlyMetProProArgGlyProGlyInserL 84  
 365 TGGCGGGAGCCCTGTGATTTGGCATGCTCCTCGGGGCCCGGAGACAGTCTC 414  
 84 eugIyGlyMetGlySerPheGlyAlaMetGlyGlnProMetSerLeuSer 100  
 415 TGGCGGGAGATGGTATCTTGGTCCATGGGACAGCCAAATGTCTCTCTCA 464  
 101 GLyGlnProProGlyThrSerGlyMetAlaProHisSerMetAlaVal 117  
 465 GGGCAGCGGCTCCTCGGACCTGGGGATGGCCCTCAGACAGATGGCTGT 514  
 117 lValSerThrAlaThyProGlnThrGlnLeuGlnLeuGlnAlaValAla 134  
 515 CGTGTCTACGGCAACGTCACAGACCCAGCTGACGCTCCAGAGAGTGGCGC 564  
 134 eugIingIingIingIingIingIingIingIingIingIingIingIingI 150  
 565 TGCAGCGACAGACAGCAACAGAGAGATTCCAGCAGCAGACAGCCGCGC 614  
 151 LeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 167  
 615 CTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 664  
 167 nSerAlaMetGlnGlnGlnPheGlnAlaValAlaGlnGlnGlnGlnGln 184  
 665 GAGTGCCATGCAGACACAGTTCAGAGCAAGTACAGCAGCAGCAGCAGCAGC 714  
 184 eugIingIingIingIingIingIingIingIingIingIingIingIingI 200  
 715 TCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 764  
 201 AsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 217  
 765 AATCAGCAACAGATACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 814  
 217 nLeuGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 234  
 815 GCTGACAGCTCCAAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 864  
 234 IngIingIingIingIingIingIingIingIingIingIingIingIing 250  
 865 AGCAGCGGGCTTTGGAGCCACCCCAATTCAGCAGCAGCAGCAGCAGCAG 914  
 251 GlnProGlnProProFoserGlnAlaLeuProGlnGlnLeuGlnGlnGln 267  
 915 CAGCAGACAGCTCCGCTCCAGGCTCTGCCACAGCAGCAGCAGCAGCAG 964  
 267 tHisHisThrGlnHisHisGlnProProGlnProGlnProGlnProPro 284  
 965 GCATCAACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1014  
 284 aAlaGlnAsnGlnProSerGlnLeuProGlnProGlnProGlnProGlnPro 300  
 1015 TTGCTCAGAACCAACATCACAACCTCCGCGCAGAGTCGCGACAGCCAGCT 1064  
 301 LeuValSerGlnAlaGlnAlaLeuProGlyGlnMetLeuTyrThrGlnPr 317  
 1065 TTGGTGTGACGGGCGGAGGCTCTCCCTGGACAATGTGTATACCCCAACC 1114  
 317 opTroleuLysPheValArgAlaProMetValValGlnGlnProProValG 334  
 1115 ACCACTGAAATTTGTCGAGCTCCGATGTGTGTGTCAGCAGCAGCCCGCAGTGC 1164

334 lInProGlnValGlnGlnGlnGlnThrAlaValGlnThrAlaGlnAlaAla 350  
 1165 AGCCCGAGGTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1214  
 351 GlnMetValAlaProGlyValGlnValSerGlnSerSerLeuProMetLe 367  
 1215 CAGATGCTGGCTCCGAGCTCCAGGTCCAGCCAGCAGCAGCAGCAGCAGCT 1264  
 367 uSerSerProSerProGlyGlnGlnValGlnThrProGlnSerMetProp 384  
 1265 GTCCTGCGCGTCACCGGCGCAGCAGGTGACAGCAGCAGCAGCAGTGCATG 1314  
 384 roProProGlnProSerProGlnProGlyGlnProSerSerGlnProsn 400  
 1315 CTCGCCCGAGCGCTGCCGCGCAGCCGCGCAGCCAGCTCAGCAGCCCAAC 1364  
 401 SerAsnValSerSerGlyProAlaProSerProSerSerPheLeuProSe 417  
 1365 TCCACAGTACGCTTGCGCTCCGCTCCCACTCCCAATAGCTTCTCCAG 1414  
 417 rProSerProGlnProSerGlnSerProValThrAlaArgThrProGlnA 434  
 1415 CCCCTACCGCAGCCCTCCAGAGCCAGTGACGGCGGAGCCGCCACAGA 1464  
 434 snPheSerValProSerProGlyProLeuAsnThrProValAsnProSer 450  
 1465 ACTTCAGTCTCCCTCACCTGGACCTTTAAACACACCTGTGAACCCAGC 1514  
 451 SerValMetSerProAlaGlySerSerGlnAlaGlnGlnGlnGlnTyrLe 467  
 1515 TCTGTGATAGCCCGCAGCTGGCTCCAGCCAGGCTGAGGAGCAGCAGTACT 1564  
 467 uAspLysLeuLysGlnLeuSerLysTyrTlIeGlnProLeuArgTyrMetI 484  
 1565 GGACAAAGCTGAAAGCAGCTGTGCAATGATCATGAGCCCTCGGCCGCAATG 1614  
 484 lAsnLysTlIeAspLysAsnGlnuAspArgLysLysAspLeuSerLysMet 500  
 1615 TCACACAGATGCACAGAACAGACAGACAGACAGACAGACAGACAGTGAATG 1664  
 501 LysSerLeuLeuAspTlIeLeuThrAspProSerLysArgCysProLeuL 517  
 1665 AAGAGCTTCTGACATTTGACAGACCCCTGGAAGCGGTGTCCCTGAA 1714  
 517 sThrLeuGlnLysCysGlnIleAlaLeuGlnLysLeuLysAsnAspMetA 534  
 1715 GACCTTGCAAAAGTGTGAGATCGCCCTGGAGAAACTCAAGATGACATGC 1764  
 534 rGcysProLeuProHisArgProAlaGcysHisArgProAsnSerSerThr 550  
 1765 GGTGCCCACTCCCGCAGCCGCGCCGCTGCCACCGCAACAGAGTAC 1814  
 551 TyrAlaSerArgSerTyrMetProSerTyrProThrSerAlaHisLeuSe 567  
 1815 TATGCCAGCGGCTCCTGGATGCGGTCGTCGCAACATCGCTCACTGCTC 1864  
 567 rSerThrIleProCysThrAlaHisSerPheGlnPro 579  
 1865 TTCAACCATTCCTGTATCCGCACATTCGTTCAGCCA 1901  
 seq\_name: gb\_pr:AF328769  
 seg\_documentation\_block: 3170 bp mRNA PRI 26-JUN-2001  
 LOCUS AF328769  
 DEFINITION Homo sapiens PC2 glutamine-rich-associated protein (PCOAP) mRNA,  
 complete cds.  
 ACCESSION AF328769  
 VERSION AF328769.1 GI:14276856  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;











```

|||||
379 GGGCAACACCCCTGGAACTCTGGAAATGCGCCCTCATGCGATGGCTGT 428
117 lvalserthrAlaThrProGlnThrGlnLeuGlnLeuGlnValAlaL 134
|||||
429 GGTTGTACAAACAACCTCAAGACTGAGTGCAGCTCCAGCAAGTGGCAT 478
134 eugGlnGlnGlnGlnGlnGlnGlnPheGln.....GlnGlnGlnGln 148
|||||
479 TGCAGCAACAGCAGCAGCGGCAACCAACAGCAATTCAGCAACAGCAG 528
149 AlAlAlLeu.....GlnGlnGlnGlnGlnGlnGlnGlnGlnGln 161
|||||
529 GCAGCACTGAGCAACAGCAGCAGCAACAGCAGCAGCAACAGCAGCA 578
161 nGlnPheGlnAlaGlnGlnSerAlaMetGlnGlnGlnPheGlnAlaVal 178
579 GCAGTTCAGGCAACAGCAATGCTGAGCAACAGTTCAGCAAGCAGTAG 628
178 alGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnHisLeu 194
629 TGCAGCAGCAGCAG.....CTTCAGCAGCAGCAGCAGCAGCAGCTG 672
195 lIleLysLeuHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 211
673 ATTAAGTTGATCATCAAGCCAGCAACAA...CAGATACACAGCAGCA 719
211 nLeuGlnAlaArgLeaGlnLeuGlnLeuGlnGlnGlnGlnGlnGlnGln 228
720 ACTGCAGAGGATGGCACAGTTGCAGCTG.....CAGG 751
228 lGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 244
752 AACAGCAACAGCAGCAACAGCAGCAGCTTGCAGGCGCCAGCAACATG 801
245 GlnGlnProMetGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 261
802 CAGCAGCATCATGCGCAGCAGCAGCAGCAGCTCCCTTCAGGCGCTAC 851
261 oGlnGlnLeuGlnGlnMetHisHisThrGlnHisHisGlnProProG 278
852 CCAGCAGCTGTACAGCTGCATCATCCAGCATCCAGCAGCAGCAGCTC 901
278 lnrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 294
902 AGGCTCAGCAGTCCCTCATTTGCTCAAAACCAACACACAGATCCACCA 951
295 GlnSerGlnThrGlnProLeuValSerGlnAlaGlnAlaLeuProGly 311
952 CAGTCACAGAGCCAGCTTTGGTGTCTCAGAGCAGCAAGCCCTTCTG 1001
311 nMetLeuTyr...ThrGlnProLeuLysPheValArgAlaProMetV 327
1002 GATGCTGTATGCTGCCACAGCAGCTGAATTTGTCGTCGTCGATGG 1051
327 alValGlnGlnProProValGlnProGlnValGln..... 338
1052 TGGTCAGAGCCGCGCAGAGTGCAGCCCGAGGTGCAAGGTCAGAGCC 1101
339 ...GlnGlnGlnThrAlaValGlnThrAlaGlnAlaAlaGlnMetVal 1354
1102 GTGCAAGCCGCGCAGCAGCAGTGCAGCAGCAGTCTGCCAGATGTA 1151
354 aproGlyVal..... 357
1152 TCCGCGGTCCAGATGATGCTCAAGCCTTGCCCAAGCGGATGCAGC 1201
357 ..... 357
1202 TAAAGACCCGGTCCCGCCACCTCCACCATGTCTGCGCCCTCAGC 1251
358 .....GlnValSerGlnSerLe 364
|||||

```

```

1252 TCACATCTCTTTGGGCGAGCAGCCCAACACAGGTCAAGCCAAACAGCCT 1301
364 uProMetLeuSerSerProSerProGlyGlnGlnValGlnThrProGln 381
1302 CACCATGCTGCTTCACCGTCAGCAGGCGCAGCAGTGCAGACCCCAAGT 1351
381 erMetProProProGlnProSerProGlnProGlyGlnProSerSer 397
1352 CGATCCCACTCCCGCAGCAGCGTCCCAACAACCTGGC.....TCA 1392
398 GlnProAsnSerAsnValSerSerGlyProAlaProSerProSerSer 414
1393 CAGCCCAACTCTAATGTACGCTCCGCGCTGCCCATCTCCACAGCTT 1442
414 eleuProSerProSerProGlnProSerGlnSerProValThrAlaArg 431
1443 CTGCTGACCCCTTACACAGCTTCTCAGAGCCGGGTGACAGCAGCA 1492
431 hrProGlnAsnPheSerValProSerProGlyProLeuAsnThrPro 447
1493 CCCCACAGACTTCAGCGTCTCTCCCTGAGACCTTTAAACACCCCTGTG 1542
448 AsnProSerSerValMetSerProAlaGlySerSerGlnAlaGlnGln 464
1543 AACCCAGCTGTGTCATGAGCCAGCTGCTGCTGAGCAGCTGAGAGCA 1592
464 nGlnIlyLeuAspLysLeuLysGlnLeuSerLysTyrTrlIleGlnProLeu 481
1593 GCAGTACCTGGCAAGCTGAAAGCAATTTGCAAGTACATCGAGCCCTGC 1642
481 rArgMetLleAsnLysLleAspLysAsnGluAspArgLysAspLeu 497
1643 GACGATGATCAACAAGATCGCAAGATGAAAGCAAGAAAAAGCATT 1692
498 SerLysMetLysSerLeuLeuAspLleLeuThrAspProSerLysArg 514
1693 AGTAAGATGAAGAGCTGCTGTCATCTCAACCGACCCCTGGAAGAGGTG 1742
514 sproLeuLysThrLeuGlnLysGlnLleAlaLeuGlnLysLeuLys 531
1743 TCCCTCGAAGACCTGCAAAAGTGAATGCTGCTGGAAGAGCTCAAGA 1792
531 snAspMetArgCysProLeuProHisArgProArgCysHisArgPro 547
1793 ATGACATGGCAGTCCCGACACCCCAACCCCGCAGTCTTCCACACAA 1842
547 nSerSerThrTyrAlaSerArgSerTrpMetProSerTrpProHis 564
1843 CAGCAGAGACTGCGCAACCACTCTAGATGCAGTCTGGCCACATCCG 1892
564 lAhHisLeuSerSerThrIleProCysThrAlaHisSerPheGlnPro 579
1893 TTCACCTGCTTCAACCAATTCCTGTACCGCAGCATTTGTGCCAGCA 1939
seq_name: gb_HSU80745
seq_documentation block:
LOCUS HSU80745 1048 bp mRNA PRI 18-DEC-1997
DEFINITION Homo sapiens CTG7a mRNA, partial cds.
ACCESSION U80745
VERSION U80745.1 GI:2565064
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1048)
Margolis R.L., Abraham M.R., Gatchell S.B., Li S.H., Kidwai A.S.,
Breschel T.S., Stine O.C., Callahan C., McIntis M.G. and Ross C.A.
cDNAs with long CAG trinucleotide repeats from human brain
Hum. Genet. 100 (1), 114-122 (1997)
TITLE JOURNAL MEDLINE
97369492
2 (bases 1 to 1048)
REFERENCE

```











Center project name: H\_NH0803A13

Sequencing vector: M13; 38%

Chemistry: Dye-primer ET; 2% of sequencing vector; plasmid; 42%

Chemistry: Dye-terminator Big Dye; 58% of reads  
Assembly program: Phrap; version 0.990319

Consensus quality: 153264 bases at least Q40  
Consensus coverage: 156105 bases at least Q40

Consensus quality: 157587 bases at least Q20

```
insert size: 108000; agarose-1p
Insert size: 169020; sum-of-con
```

quality coverage: 9.85 in Q20 bases; agarose-fp  
quality coverage: 9.86 in Q20 bases; sum-of-con

\* NOTE: This is a 'working draft' document. It contains errors and omissions.

```

* NOTE: This is a working draft sequence. It currently
* consists of 6 conligs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the conligs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

FEATURES			
*	107517	conf: 0.48799	bp in length
*	107515	gap of	unknown length
*	107416	conf: 0.45392	bp in length
*	61925	gap of	unknown length
*	39186	conf: 0.22739	bp in length
*	39086	gap of	unknown length
*	33085	conf: 0.21398	bp in length
*	17688	conf: 0.39085	bp in length
*	9272	gap of	8315 bp in length
*	9173	conf: 0.9172	bp in length
*	1	conf: 0.9172	bp in length







```

403 ..... 403
657 AGGCTGGCCTGGGGTGGCCCTCCACCTGGCCCTGGAGCTGGCCCTG 706
404 .....Ser..SerGlyPro 407
707 CCTGGACCCCGGCTGCATGTTTCTCTCACTGGCTGCACGCTGGCCCT 756
408 AlaProSerProSerPheLeuProSerProSerProGlnProSerG1 424
757 GCCCATCTCCAGTACCTCTCTGCCAGCCCTCACCGCAGCCCTCCCA 806
424 nSerProValThrAlaArgThrProGlnAsnPheSerValProSerProG 441
807 GAGCCAGTGAAGGCGCGAGCCCGACAGACTTCACTGCTCCCTCACCTG 856
441 LProLeuAsnThrPro..Val..... 447
857 GACCTTTAAACACACACTGTGTAAGTTGGGCTGAGGCTGCTAAGGTACCTCC 906
447 ..... 447
907 TCACCTTTATGAGGCTTCAGCTCATCTGGGTGCGAGCTGGGGGCC 956
447 ..... 447
957 TCAGAGCTCAAGTTCCACCCCGAGGGGTGAGGGCTGGCCCTCACCCGC 1006
448 .....LysProSerSerValMetSerProAlaGlySer 459
1007 CTGTGCTCCTGCAGTGACCCAGCTCTGTCTATGAGCCAGCTGGCTCGAG 1056
459 TGIAlaGlnGlnGlnGlnTyrLeuAspPheLeuGlnLeuSerLysT 476
1057 CCAGCTGAGGAGCAGCAGTACTGAGCAACTGAGAGAGCTGTGCAAGT 1106
476 TyrLeuProLeuAsnGlnMetLeuAsnLysLysLysLysLysLysLys 492
1107 ACATGAGACCCCTGCCCGCATGATCAACAAGATCGACACAGACAGAC 1156
493 ArgLysLysAspLeuSerLysMetLysSerLeuLeuAspLysLysLys 509
1157 AGAATAAAGAGACTGTGTAGATGAAGAGCCCTGTGACATTTGACAGA 1206
509 ProSer..... 511
1207 CCCTCGAAGGGGTGTGCTTTGCCACAGCCACGAGGGTCCACAAAGG 1256
511 ..... 511
1257 CACAGATAGCCAGCCATGATGGGCACTTGATGATGATGGGTTAAC 1306
511 ..... 511
1307 AAAGGACACAGGAGCTCTTTGAGACCTGGCCAGAGGCTTCACAGGCTCC 1356
511 ..... 511
1357 ACTGTGTGTGTGTGTGGGGCTTCAAGCCAGGCTTCATCTTGGCCATGC 1406
511 ..... 511
1407 CCAGCTTTGGCCATTCCTGCTGCTGCTGTGTGGCCCTCATGCTGGGCCA 1456
511 ..... 511
1457 TCACAGCCAGACCTCATCCAGTCAGCAGCAGAGGGCCAGCTTGCAGTGG 1506
511 ..... 511
1507 CCTGACCATGAGCTGGCCAGATGGGCTGAGCCTGACCTGAGAGTTCTGC 1556
511 ..... 511
1557 CCCTGACTTGTGTATGACCATGAGGAGGAGGAGCTGATCTGATTAACCCA 1606
511 ..... 511
1607 TTGTATGCCACCTGTGTGTATGATAAATAAACAAGGTGGCGCTGTGTG 1656
511 ..... 511
1657 AGAAGCAGTGTGAAGGAGGAGGCTGTGGCACAGGCTGTGTGAGATGA 1706
511 ..... 511
1707 GAGCTGGGAGTACCTCCAACTGCATGGGATTTGGCATATACCTCCCT 1756
511 ..... 511
1757 TTACCTCAACTGAGGCTATCCCCAACCCCTGTGTAGAGGCTCTCT 1806
511 ..... 511
1807 GTTCTGCTGTCTGTATAGATGACAGTCACTATTGGGTGTCTCCAGG 1856
511 ..... 511
1857 TCTTCATGGGTGATCCTTTGTTAGGGTTATAGGATGATGGGCGAGGG 1906
511 ..... 511
1907 ACTGGCTACATGACAAATGAGGATTCACAGGCTGATCAAGGGCTGGGC 1956
511 ..... 511
1957 ACTGTCCCTTCTCTTACCATTCACCTCAGCTGCGCAACGAGGAGAACG 2006
511 ..... 511
2007 AGCTCCTGGGGGTGAGGGCTGCCCTGCTTCAACTGTGGGTGAGAACAG 2056
511 ..... 511
2057 GTGGTCACTGTAGAGGCTCCTGCAGAGGTTTGTATGCTGAGGCCCA 2106
511 ..... 511
2107 GTGGCAGTAGTTGGATCCTGAGCCTGAGAACTGCATGTCTTCCACTCT 2156
512 .....Lys..ArgCysProLeuLysThrLeuGln 521
2157 TTCAGCCCAAGTCTCTCTCTGTGCGAGGTGTCCCTGAAGACTTGCAAA 2206
521 yscysGlnLeuAlaLeuGlnLysLysLysLysLysLysLysLysLys 534
2207 AGGTGTAGATGCCCTCGGAGAACTCAAGAAATGACATGGCGGTGTGAGT 2256
534 ..... 534
2257 GGGATCCAGACACCCCTAGGGGAACAGGGCTCTCTTAAGACTCTCTGG 2306
534 ..... 534
2307 GAGTGGCGGTGAGAGGGCTTCAGAGTCAGGATGTGGCGGGGGCGG 2356
534 ..... 534
2357 GCCTGACCTTGGACCCCTGCCAGAGGCTTCTGGCCAGGTCTGTGTGC 2406
534 ..... 534
2407 TGGGTGGAACATGGAGAAAGTCAACCTCTGTCTAGGCGCCCGTGGGTG 2456
534 ..... 534

```



















511 ..... 511  
 31894 TGTATGAAATATGAACAAGTGGCGCTGGTGAGAACAGAGTGGAAAGCA 31943  
 511 ..... 511  
 31944 GGGCTGTGGCCACAGGCTGCTGTGAGAGATCAGAGCTGGGCGAGTACCTC 31993  
 511 ..... 511  
 31994 CAACATGACATGGCATTGGCATATACCTCCCTTACCTCAGCTCAGAGCT 32043  
 511 ..... 511  
 32044 ATCCCCCAACCCCTGTCTAGAAAGGCTCTCTGTCTGTCTGTCTGTGATA 32093  
 511 ..... 511  
 32094 GAATGCCAGTCATATGGGTGCTCTCAGGCTTATGGGTGATCTCT 32143  
 511 ..... 511  
 32144 TTGTTAGGTTATAGGATGAGTGGGCGAGGAGCTGGCTTACATGACAT 32193  
 511 ..... 511  
 32194 GAGGATTCACAGAGCTGATCAGAGGCGCTGCGCACTGCTCTCTCTTGAC 32243  
 511 ..... 511  
 32244 CCATCACTCACTCTGCCAACAGGAGCAAGAGCTCTCTGGGTTGAGGG 32293  
 511 ..... 511  
 32294 CTGCCTGCCCTTCACTCTGGGTCTGGGAACAGCTGTCAACGTGTAGAG 32343  
 511 ..... 511  
 32344 CTCCTGACAGAGGTTCTGATGGCTGAGGCCAGTGCAGTACTTGGATC 32393  
 511 ..... 511  
 32394 CTGAGCCTGAGACCTGCTCTTCCACTCTTCCAGCCCAAGCTCTCTC 32443  
 512 ..Lys..ArgCysProLeuLysThrLeuGlnLysCysGlnIleAlaLeuG 527  
 32444 TCCTGTCAGGTGTCCTCCCTGAAAGACCTTGCAAAAGTGTGAGATCGCCCTGG 32493  
 527 LulysLeuLysAsnAspMet Arg ..... 534  
 32494 AGAACTCAAGATGACATGGCGGTGTGATGGATGCCACACACCCCT 32543  
 534 ..... 534  
 32544 AGGGAAACACAGGCTCTCTTAAGAGCTCTGGAGTGTGTGAGAGGCG 32593  
 534 ..... 534  
 32594 CTTCAAGGTGACAGGATCTGGGCGGGGCGGCGCTGACCTTGGACCTTG 32643  
 534 ..... 534  
 32644 CCCACAGAGCTTCTCTGCGCAGCTCTGCTGTGCTGGGTGGGAACATGGAG 32693  
 534 ..... 534  
 32694 AAGTCACTCTGTCTAGAGCCCGCGGTGGTCTCTTACCACTTTCATCA 32743  
 534 ..... 534  
 32744 GAGAGGCGAGGAGCTGTAGGAGATAGGCTGTGGGGTGGGGGATTTAT 32793

534 ..... 534  
 32794 CCAGAGTACGCCACATTTCTGCAACAAGGTGGAGGCTGTGAGAGAG 32843  
 534 ..... 534  
 32844 CCTCGGAATCTGACTGTGATGATGACATGGCGCTGGGTGTGAAAGCCCG 32893  
 534 ..... 534  
 32894 CTAAATGGGGAACCTCGGGTCCCGGGCTGCTGACCGTCCATCTCTGC 32943  
 535 ...CysProLeuProHisArgProArgCysHisArgProAsnSerThr 550  
 32944 TCCAGGCCACCTCCCGCCCGCGGTGCGACACGCAAAAGCAGTAC 32993  
 550 rTyraIAserArgSerTrpMetProSerTrpProThrSerAlaHisLeu 567  
 32994 CTATGCCAGCGCTCTGATGCCGCTGCTGCGCCACATCCGCTCACCTGT 33043  
 567 eSerThrIleProCysThrAlaHisSerPheGlnPro 579  
 33044 CTTCAACCATTTCCCTGTACCGCACATTTCTTCCAGCCA 33081  
 seq\_name: gb\_hhg:AL358856

seq\_documentation block:  
 LOCUS AL358856 157080 bp DNA HTG 10-JUN-2001  
 DEFINITION Homo sapiens chromosome 1 clone RP11-4623, \*\*\* SEQUENCING IN  
 PROGRESS \*\*\*, 12 unordered pieces.  
 ACCESSION AL358856  
 VERSION AL358856.3 GI:9863846  
 KEYWORDS HTG: HTGS\_PHASE1; HTGS\_CANCELLED.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
 AUTHORS Mclay, K.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 requests: clonerequest@sanger.ac.uk  
 On Aug 22, 2000 this sequence version replaced gi:9214244.  
 COMMENT  
 ----- Genome Center  
 Center: Sanger Centre  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquery@sanger.ac.uk  
 ----- Project Information  
 Center project name: BA4G23  
 ----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Sequencing vector: plasmid; L08752; 100% of reads  
 Chemistry: 15367 bases at least Q40  
 Consensus quality: 15799 bases at least Q30  
 Consensus quality: 154743 bases at least Q20  
 Insert size: 155980; sum-of-contigs  
 Insert size: 165281; 6.3% error; agarose-fp  
 Quality coverage: 5.20x in Q20 bases; sum-of-contigs Quality  
 coverage: 4.91x in Q20 bases; agarose-fp

NOTE: This is a 'working draft' sequence. It currently  
 consists of 12 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.  
 \* 1 3463: contig of 3463 bp in length  
 \* 3464 3563: gap of 100 bp



```

* 3564 13200: contig of 9637 bp in length
* 13201 13300: gap of 100 bp in length
* 13301 59189: contig of 45889 bp in length
* 59190 59289: gap of 100 bp in length
* 59290 85824: contig of 26535 bp in length
* 85825 85924: gap of 100 bp in length
* 85925 88142: contig of 2218 bp in length
* 88143 88242: gap of 100 bp in length
* 88243 95133: contig of 6791 bp in length
* 95134 100091: gap of 100 bp in length
* 100092 100191: gap of 100 bp in length
* 100192 117326: contig of 17135 bp in length
* 117327 117426: gap of 100 bp in length
* 117427 119946: contig of 2520 bp in length
* 119947 120046: gap of 100 bp in length
* 120047 130509: contig of 10463 bp in length
* 130510 130609: gap of 100 bp in length
* 130610 139309: contig of 8700 bp in length
* 139310 139409: gap of 100 bp in length
* 139410 157080: contig of 17671 bp in length.
Location/Qualifiers
1. 157080
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-4623"
/clone_1lb="RPCT-11.1"
1. 33463
/note="assembly-fragment:00230
clone_end:SP6
vector_side:left"
3564. 13200
/note="assembly-fragment:00057
fragment_chain:1"
13301. 59189
/note="assembly-fragment:01204
fragment_chain:1"
59230. 85824
/note="assembly-fragment:01664
fragment_chain:1"
85925. 88142
/note="assembly-fragment:00505
fragment_chain:1"
88243. 95033
/note="assembly-fragment:01296
fragment_chain:2"
95134. 100091
/note="assembly-fragment:01408
fragment_chain:2"
100192. 117326
/note="assembly-fragment:00869
fragment_chain:2"
117327. 119946
/note="assembly-fragment:00628
fragment_chain:3"
120047. 130509
/note="assembly-fragment:01620
fragment_chain:3"
130610. 139309
/note="assembly-fragment:01199"
139410. 157080
/note="assembly-fragment:01779"

```

```

FEATURES
source
misc_feature
/note="assembly-fragment:00230
clone_end:SP6
vector_side:left"
3564. 13200
/note="assembly-fragment:00057
fragment_chain:1"
13301. 59189
/note="assembly-fragment:01204
fragment_chain:1"
59230. 85824
/note="assembly-fragment:01664
fragment_chain:1"
85925. 88142
/note="assembly-fragment:00505
fragment_chain:1"
88243. 95033
/note="assembly-fragment:01296
fragment_chain:2"
95134. 100091
/note="assembly-fragment:01408
fragment_chain:2"
100192. 117326
/note="assembly-fragment:00869
fragment_chain:2"
117327. 119946
/note="assembly-fragment:00628
fragment_chain:3"
120047. 130509
/note="assembly-fragment:01620
fragment_chain:3"
130610. 139309
/note="assembly-fragment:01199"
139410. 157080
/note="assembly-fragment:01779"
BASE COUNT 37302 a 35841 c 39467 g 39369 t 1101 others
ORIGIN

```

```

alignment_scores:
Quality: 705 50
Ratio: 2.486
Percent Similarity: 27.805
Percent Identity: 23.415
alignment_block:

```

```

US-09-668-119-3 x AL358856/rev ..
Align seq 1/1 to reverse of: AL358856 from: 1 to: 157080
39 HispEarGasp1LeHisAnLysSergL..... 49
|||||
77750 TATTATAGACTAATCATGATAGGAAATCCCAAAATCAACACAAATGCTA 77701
50 .....AlaserVaserAspPrometAnAlaLeuGlnSerL 62
|||||
77700 CTTCGATTCCTTCTCTCTGCTGTAATGGAAGCTTCAGATT.... 77655
62 eurtHrgLysGlyProAlaLagLAlaLagLLeuGlyMetPropPArg 78
|||||
77654 .....CCAGCGGA...GCCGTGGCACTGGGGCTTTG... 77625
79 GlyProGlyLysLeuGlyMetGlySerPheGlyAlaMetGlyL 95
|||||
77624 .....CTGGCTGTTTGTCCCTGCTGTGCA 77599
95 nProMetSerLeuSerGlyGlnProProGlyThrSerGlyMetAlaP 112
|||||
77598 GCCCTGCAGCGTTTCTGGAATCTGCCCTGTGAGCTGACTGC... 77556
112 roHisSerMetAlaValSerThrAlaThrProGlnThrGlnLeuGln 128
|||||
77555 .....GACTCGTCTTTCACGCCACGCTGCAG 77526
129 LeuGlnGlnValAlaLeuGlnGlnGlnGlnGlnGlnGlnPheGlnL 145
|||||
77525 CTCACAGCAGGTGGCGCTGCACGACGACGACGACGACGACGACGACG 77476
145 ngInGlnGlnAlaAlaLeuGlnGlnGlnGlnGlnGlnGlnGlnGln 162
|||||
77475 GCAGCGACGAGCGCGCTCTACAGAGCAGCAGCAGCAGCAGCAGCAGC 77426
162 InPheGlnAlaGlnGlnSerAlaMetGlnGlnGlnPheGlnAlaVal 178
|||||
77425 AGTTCAGGCTCAGCAGAGTGCATGACGACGACGACGACGACGACG 77376
179 GlnGlnGlnGlnGlnLeuGlnGlnGlnGlnGlnGlnGlnHisLeuL 195
|||||
77375 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 77326
195 eLysLeuHisHisGlnAsnGlnGln..... 203
|||||
77325 TAAATTGCATCATCAAAATCAGCAACAGGTACAGGTCCCTGTTCTGC 77276
203 ..... 203
77275 TCTTGGCTCCCTCTGCAGCGTGCAGCCCTGGCTGCATGACCAAT 77226
203 ..... 203
77225 GCTGGTGTCAGTCCCGGAGCCAGCGAGGTTCTTGTACAGCTGTGCC 77176
203 ..... 203
77175 TGCCCTTTCCATGGGCTTCTCAAAAAGCTTGCGACAAGCGCGGCCAG 77126
203 ..... 203
77125 TGGCTCATGCTGTAAATCCAGCACTTTGGAGGCCAAGCAGCGGATC 77076
203 ..... 203
77075 ATCAGAGATTCAAGACAGCGCTGGCCAGCATGTGGAATCCCATCTCTAC 77026
203 ..... 203
77025 TAAATAATACAAAATTAGCTGGACATGTGGGTACACTAATATCCCA 76976
203 ..... 203

```



76975 GCTATTGGAGGCTGAGCAGAGAAATCGTAGAACTGGAGCGGAG 76926  
203 ..... 203  
76925 GTTGCAGTACCGAGATGCGCGCTGCGCTCCAGCTGGCAGCAGG 76876  
203 ..... 203  
76875 AGACTGTCTCAAAAAATAAATAAAGAAAGCAAAAGTCTGG 76826  
203 ..... 203  
76825 GACAGCAGAGAAACAGGCTGACTACTAGAAAGTTCTGTATGCA 76776  
203 ..... 203  
76775 GCCCTTCATTGGGCTTGGGCATCTGGCGTGGGCTCCTGCTAGTTG 76726  
203 ..... 203  
76725 AGGGCTGCGAGGAAITCAAAACAAGTAGCTGAAGCAAGTAGAGC 76676  
203 ..... 203  
76675 GAACACTCTGAGGCTTGGACTACTGAGCCATTGGGCCATTGGCT 76626  
203 ..... 203  
76625 GAGACACATGTGCTCCTTTACCTGTCTTGGGTTGGGCTCA 76576  
203 ..... 203  
76575 GTGTAGTCGACCTCAGCTGGTGGTTCATGAGGTGATGTCAG 76526  
203 ..... 203  
76525 TTAGCAGACTGCTTCAGAGTTGTTCACTCCATGGTTCCATAG 76476  
203 ..... 203  
76475 AGAGTACAGTGTAAACATCTGCCCCACAGCAGGCTGTGGTGGGA 76426  
203 ..... 203  
76425 GAGGAAAGCCCAAGGCCCCACCCTCCACCCCTCAGGCTCAG 76376  
203 ..... 203  
76375 CTAGTGGGCTTCTGTGATCTGGCAACCCCTCCAGAGCTGACAGTG 76326  
203 ..... 203  
76325 CATCTGACGCCACATTTCTCTCTCTGCTGACAGACGACCTTGACCG 76276  
203 ..... 203  
76275 TCACCGTAGGCTCATGCCACTGACCTATAGGGCAAGCAGATGTT 76226  
203 ..... 203  
76225 GCAGAGCTTCAGAAATCTGCAGTTGCCCGAGTTTATCTTACCTTCA 76176  
203 ..... 203  
76175 TGGAGCTTACCTATCTCTGTCACTTTTGTGAAGTTCTTACGTATGC 76126  
203 ..... 203  
76125 CCTTGAACCATCTAGTGAATGTAGACAGAGCCAGCAGACACTT 76076  
203 ..... 203

76075 GGTGCCAGGCAACACTTGTGCTCCCTCCAGTGTACGAAAGCAGCTGT 76026  
203 ..... 203  
76025 GTTGGCTCGGGGGGCCAGAGCCTTTCTGTGGAGATGCCCCCTCCC 75976  
203 ..... 203  
75975 GTGTAGTAGAGGTGACGCTGTCAAGCCAGAGAGGCTTTTTTTTTT 75926  
203 ..... 203  
75925 TTATTGAGACTTTTTTTGCTTACTCTGTGTGCCAGCTGAGATGCA 75876  
203 ..... 203  
75875 GTGTGTATCTCGGCTCAGTCAACCTGACCCACAGTTTAAGTAT 75826  
203 ..... 203  
75825 TCTCTGCTCAGCTCCCAAGTACGTGGATTAAGGTGAGCCACCA 75776  
203 ..... 203  
75775 TGCTGGCTAATTTTGTATATATAGTAGAGCGGGTTGCGCAGTTG 75726  
203 ..... 203  
75725 TCCAGCTGTCTCGAATCTGACCTCAGGTATGATGCCCGCTCGGC 75676  
203 ..... 203  
75675 TCCCAAGTCTGGGATTACAGCGGTGAGCCACTGCACCCAGCAGAA 75626  
203 ..... 203  
75625 GCTTCTGTATGAGCTTCGACATGCCACCTGGGCTTCCTGAGGCCAGA 75576  
203 ..... 203  
75575 CTGTACAGGAGAGACTGTGCCACAGAGCAGATGATGATACCGAG 75526  
203 ..... 203  
75525 TGCTGCTGTCTGTCTAGATACAGCAGCAACACACACTCAGCG 75476  
203 ..... 203  
75475 AATGACACAGCTGACCTCCAAACACAGCAGCAGCAGCAGCAGCAGC 75426  
203 ..... 203  
75425 AGCAGCAGCAGCAGCGCTTTCAGAGCCAGCAGCAGCAGCAGCAGC 75376  
203 ..... 203  
75375 CCGATGACAGCAGCAGCGCTCCGCTCCAGGCTCTCCAGCAGCAGCT 75326  
203 ..... 203  
75325 GACAGATGATCATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 75276  
203 ..... 203  
75275 AGCCTCAGTGTCTCAGAACCAACCACTCCCGCAGCTGAGCAG 75226  
203 ..... 203  
75225 ACCAGGCTTTGGTGTACAGCGCAAGCTCTCCGTCGCAAAATGTTGA 75176  
203 ..... 203  
75175 TACCAACACCACTGAAATTTGTG...AGTACTGTGGCCACAGTGA 75126  
203 ..... 203



```

331 roProValGlnProGlnValGlnGlnInThrAlaValGlnThrAla 347
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
75128 GCACATGACGCGCGTACCTCTGTGACAGATATTCTAGCGCTTGGCT 75079
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
348 GlnAlaIaGlnMetValAlaProGlnValGlnValSerGlnSerLe 364
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
75078 AAGATAGATATCCATTCTTCAGAGTGTGAGACTTCAGGACGCCCA 75029
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
364 uProMetLeuSerSerProSer..... 371
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
75028 CCCCTTGCCAGCCCTACCGACTTATGATGCTGCTGACGAGGAGCATGTG 74979
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
372 .....ProGlnGlnVal 376
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
74978 TTTTCATTGTTGTCTGAGAGCTTCTGACCCCGCAGAGGCTTCT 74929
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
377 GlnThrProGlnSerMetProProProProGlnProSerProGlnProG 393
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
74928 GCTGCTGCTGGGACCA.....AGTCG 74903
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
393 yGlnProSerSerGlnProAsnSerAsnValSerSeryl..... 406
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
74902 CATTCAGACAGCATGTGCTGATGACGACATTTGTGGGCCCTTCACGG 74853
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
407 .....ProAlaPro 409
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
74852 TGTGACACTTCGACATGATGACACGACATGAAAGATGTGACGCTCCG 74803
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
410 SerProSerSerPheLeuProSer 417
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
74802 TCAGTGTCTGCTGTGCTTTCACACT 74779
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

seq_name: gb_hlg:AC087802;

seq documentation block:
LOCUS AC087802 187315 bp DNA HTG 23-APR-2001
DEFINITION Mus musculus chromosome 16 clone rp23-472f15, WORKING DRAFT
SEQUENCE 4 ordered pieces.
ACCESSION AC087802.5 C1:13752732
VERSION AC087802.5
KEYWORDS HTG; HTGS-PH/SE2; HTGS-DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 187315)
Mus musculus Chromosome 16 BAC Clone rp23-472f15
2 (bases 1 to 187315)
Jiang, X., Sorg, L. and Roe, B.A.
Unpublished
Submitted (26-JAN-2001) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
On Apr 23, 2001 this sequence version replaced gi:12740049.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 7409 7408: contig of 7408 bp in length
* 7509 7508: gap of unknown length
* 17333 17332: contig of 9624 bp in length
* 17433 17432: gap of unknown length
* 93194 93194: contig of 75762 bp in length
* 93195 93194: gap of unknown length
* 93295 187315: contig of 94021 bp in length.

```

```

FEATURES
    source          Location/Qualifiers
                    1..187315
                    /organism="Mus musculus"
                    /db_xref="taxon:10090"
                    /chromosome="16"
                    /clone="rp23-472f15"
                    /clone_11b="rp23"
BASE COUNT      49420 a 44136 c 42281 g 51159 t      319 others
ORIGIN
alignment_scores:
    Quality: 679.00      Length: 457
    Ratio: 3.031      Gaps: 15
    Percent Similarity: 49.015      Percent Identity: 41.357
alignment_block:
    US-09-668-119-3 x AC087802/rev ..
Align seg 1/1 to reverse of: AC087802 from: 1 to: 187315
276 ProProGlnProGlnInProProValAlaGlnAsnGlnProSerGlnLe 292
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
55179 CCGAGGCTCCCTCCTTATACACACAGTGTGAGAGACAGGCCATTACCAT 55130
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
292 upro...ProGlnSerGlnThrGlnProLeuValSerGlnAlaVal 308
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
55129 ATCTATACACAGCCCAAGAAACCTGTTGCTGCTCCACCAA.... 55084
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
308 euProGlnGlnMetLeuTyThrGlnPro.....ProLeuLysPhe 321
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
55083 ....6GAGCAGGCTACCTACCTGCTGCTTCTTCTTACGCTTGGATAC 55039
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
322 ValArgAlaProMetValAlaGlnInProProValGlnProGlnValG 338
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
55038 .....CATGTGTATGTCTGCCAGAGGTAGCTGTGCTGCTGCTGATAA 54995
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
338 ngGlnGlnInThrAlaValGlnThrAlaGlnAlaGlnMetValAlaP 355
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
54994 AGAAATPAGGGTGGATAGATGATGATGATGATGATGATGATGATGAT 54945
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
355 roGlyVal.....GlnValSer 360
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
54944 AGGTGTGTGTGATGATGATGATGATGATGATGATGATGATGATGATG 54895
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
361 GlnSerSerLeuProMetLeuSerSerProSerProGlnGlnValG 377
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
54894 CAAAGAGGCTTACCATGCTGTGCTGATGATGATGATGATGATGATGAT 54845
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
377 nThrProGlnSerMetProProProProGlnProSerProGlnProG 394
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
54844 GACCCACAGTGTGATGATGATGATGATGATGATGATGATGATGATG 54796
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
394 InProSerSerGlnProAsnSerAsn..... 402
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
54795 .....TCACAGCCCAACTATATGTCAGTAGGCTGTGGGATGAC 54754
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
402 ..... 402
54753 COTTCCTTAGCTTCCTTACCCCATTTGGCTTAGGCTGACATGCTGT 54704
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
403 .....Val.SerSerylProAla 408
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
54703 TCTAAGCTCAGACCTGCTTTATTATCATTTGCTGTGAGTGTGGGCTTGC 54654
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
409 ProSerProSerSerPheLeuProSerProSerProGlnProSerGlnSe 425
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
54653 CCATTCCTCCAGAGCTTCTCTGCTTACCCCTTACACACAGCTTCTCAGAG 54604
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
425 rProValThrAlaArgThrProGlnAsnPheSerValProSerProGlyP 442
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
54603 CCAAGTGAAGAGAGCAGCCACAGAAATTCAGGCTTCTTCTTCCCTTGAC 54554
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```



```

REFERENCE
AUTHORS      1 (bases 1 to 228467)
TITLE        Yang, L., Hu, P., Wu, H., Prescott, A. and Roe, B.A.
JOURNAL      Mus musculus BAC Clone rp23-5117
REFERENCE    2 (bases 1 to 228467)
AUTHORS      Yang, L., Hu, P., Wu, H., Prescott, A. and Roe, B.A.
TITLE        Unpublished
JOURNAL      Direct Submission
COMMENT      Submitted (17-AUG-2000) Department of Chemistry And Biochemistry,
              The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
              On Jul 31, 2001 this sequence version replaced gi:14861748.
              * NOTE: This is a 'working draft' sequence. It currently
              * consists of 2 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              *
              * 1 90075: contig of 90075 bp in length
              * 90076 90175: gap of unknown length
              * 90176 228467: contig of 138292 bp in length.
FEATURES
source       Location/Qualifiers
              1..228467
              /organism="Mus musculus"
              /db_xref="taxon:10090"
              /chromosome="16"
              /clone="rp23-5117"
              /clone_1fb="RP23"
BASE COUNT   61424 a 52795 c 53888 g 60257 t 103 others
ORIGIN
alignment_scores:
      Quality: 679.00      Length: 457
      Ratio: 3.031      Gaps: 15
Percent Similarity: 49.015 Percent Identity: 41.357
alignment_block:
US-09-668-119-3 x AC079044/rev ..
Align seg 1/1 to reverse of: AC079044 from: 1 to: 228467
276 ProProGlnProGlnGlnProProValAlaGlnAsnGlnProSerGlnLe 292
      ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
15960 CCGAGGCTCCCTCCTTATACACCCAGCTTGAGAGGAGCCATTACCAT 51911
292 uPro...ProGlnSerGlnThrGlnProLeuValSerGlnAlaGlnAla 308
      ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51910 ATCTATACCAAGCCAGAAACCTGAATCTCTTGCTGTCCACCA... 51865
308 euProGlyGlnMetLeuTyThrGlnPro.....ProLeuTyrSph 321
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
51864 ....GGAGCCAGGCTACCTACTGCGCTTCTCTAGCCCTTGAGATAC 51820
322 ValArgAlaProMetValValGlnGlnProProValGlnProGlnValG 338
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
51819 .....CATGTATTATGTGTCGCCAGAGTAGGCTTGCTGCTGAGTAA 51776
338 ngGlnGlnGlnThrAlaValGlnThrAlaGlnAlaGlnMetValAlaP 355
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
51775 AGAATAAGAGGGTGATTAAGATGGAGTACAGAGAGTGAGGAGTTTGAC 51726
355 roGlyVal.....GlnValSer 360
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51725 AGGGTGTGTAGCATGAGACTGGCTAAATCTCTCTCACTTCCAGCTACG 51676
361 GlnSerLeuProMetLeuSerSerProSerProGlyGlnGlnValG 377
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51675 CAAGACAGCCTCACAGCTGTCTCACCCTGACAGGCCAGCAGAGTACA 51626
377 nThProGlnSerMetProProProGlnProSerProGlnProGly 394
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```



```

51625 GACCCACAGTGTGATGCCACCTCCGCCACAGCGTCCGCCACACACTGAC. 51577
394 InProSerSerGlnProAsnSerAsn..... 402
|||||
51576 .....TCACAGCCCAACTTAATGTCAGTAGTGTGCTGGATGAC 51535
402 ..... 402
51534 CCTTCCTTAGCCTCCTTACCCCATTTGGCCTTAGGGCTGGCATCGTGT 51485
403 .....Val.SerSerGlyProAla 408
|||||
51484 TCTAGAGCTAGACACGCTTTTATTCACCTTGGCTGAGCTCGGGCTGACC 51435
409 ProSerProSerSerPheLeuProSerProSerProGlnProSerGlnSe 425
|||||
51434 CCATCTCCAGCAGCTCTCTGCTAGCCCTTACACACAGCCTTCACAGAG 51385
425 IProValThrAlaArgThrProGlnAsnPheSerValProSerProGlyP 442
|||||
51384 CCCAGTGCAGACGACGACCCACAGAACTTCAGCGTCTCCCTCGCTGAGC 51335
442 IProValAsnThrPro..... 446
|||||
51334 CTTTAAACACCCCTGTAAGTTGGGCGCTGGGTGAGGACGCTTACGA 51285
446 ..... 446
51284 AGGTGCTTAATCCCTTCCTCCCTCCAGAGGCTCAGATTGGTTAAAGGCT 51235
446 ..... 446
51234 CATGGAGCCCTCACAATCTCTGTGACACCTGCTCCACAGTTGAGCCAT 51185
447 .....ValAsnProSerSerValMetSer 454
|||||
51184 CCATCTAACCCACCTGTCCCTATAGTGAACCCAGCTCTGTCATGAGC 51135
455 ProAlaGlySerSerGlnAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 471
|||||
51134 CCAGTGGCTCTAGCCAGGCTGAGAGCAGCAGTACCTGAGCAAGCTGAA 51085
471 GlnLeuSerLysTyrIleGluProLeuArgArgMetIleAsnLysIleA 488
|||||
51084 GCAATTGTCAGATGATCATCGAGCCCTCGACGATGATCAACAAGATCG 51035
488 sPlyAsnGlu..... 491
|||||
51034 ACACAGATGAAAGCTAGGCTAGCCAAAGCAGATGGGCGGACACCCAA 50985
492 .....AspArgLysLysAspLeuSe 498
|||||
50984 GGCCTGTTCTTAGTGTCCCTTTCTTCTTACACAGAAAAAAGACTTAG 50935
498 IlyMetLysSerLeuLeuAspIleLeuThrAspProSerLysArg.... 513
|||||
50934 TAAAGTGAAGAGCCGTGACATCTCACCCGACCCCTCGAAGAGGTAG 50885
514 .....CysProLeuLysThrLeu..... 519
|||||
50884 TACTGGCCAGAGGCTTTGGGACAGGCTGTGTTGGCCACAGCTCACAATAG 50835
520 .....GlnLysCys 522
|||||
50834 GGTCCCAAGACTGTGTGACAGTATGATTAATCAAAAGTCCAAAGCG 50785
522 sGluIleAlaLeuGluLys.....LeuLysAsnAspMetArgCysProL 537
|||||
50784 CTCTGTGACACTGCTAGGCTTAGGCTTACAAAGTGGGTGTCAG 50735
537 euProHisArgProArgCysHisArgProAsnSerThrTyrAlaSer 553
|||||
50734 GGAACACTTACCCCTCCGAGACCTT.....TCCAGCACCCACCTTCA 50691

```

```

554 ArgSerTyrMetProSerTrpProThrSerAlaHisLeuSerSerThrIle 570
|||||
50690 AGGAGTTTGCT.....TCACACTTT 50668
570 eProCysThrAlaHisSer 576
|||||
50667 CACTGTGGCTGTCAATAGC 50649
seq_name: gp_hcrg:AL359452
seq_documentation_block:
LOCUS AL359452 64231 bp DNA HTG 13-JUN-2001
DEFINITION Homo sapiens chromosome 1 clone RP11-422P18, *** SEQUENCING IN
PROGRESS ***, 20 unordered pieces.
ACCESSION AL359452
VERSION AL359452.4 GI:9930964
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 64231)
REFERENCE
AUTHORS Plumb B.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
On Aug 27, 2000 this sequence version replaced gi:9864141.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
Project Information
Center project name: ba422p18
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: dye-terminator Big Dye; 100% of reads
Consensus quality: 52454 bases at least Q40
Consensus quality: 59753 bases at least Q30
Insert size: 6231; sum-of-contigs
Insert size: 190157; 8.2% error; agarose-fp
Quality coverage: 2.09% in Q20 bases; sum-of-contigs quality
coverage: 1.09% in Q20 bases; agarose-fp
* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 2992: contig of 2992 bp in length
* 2993 3092: gap of 100 bp
* 3093 5136: contig of 2044 bp in length
* 5137 5236: gap of 100 bp
* 5237 7850: contig of 2614 bp in length
* 7851 7950: gap of 100 bp
* 7951 9976: contig of 2026 bp in length
* 9977 10076: gap of 100 bp
* 10077 12531: contig of 2455 bp in length
* 12532 12631: gap of 100 bp
* 12632 14688: contig of 2057 bp in length
* 14689 14788: gap of 100 bp
* 14789 18056: contig of 3268 bp in length
* 18057 18156: gap of 100 bp
* 18157 21784: contig of 3628 bp in length
* 21785 21884: gap of 100 bp
* 21885 24801: contig of 2917 bp in length

```















Fri Mar 1 09:12:02 2002

us-09-668-119-3.rge

Page 24

1634 GCAGATTCCTGCAATCCTACACAGCGGTGCCCTGGAGACTTACTTA 1683  
521 yscysgluilealaleugluylsleu 529  
||||||| |||||||  
1684 AGTGTGAAAAAGCGCCTTGAGAAGATG 1709







PA (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,  
PI Wang J, Wang Z, Wähman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Woodrich R, Drmanac RT;  
XX  
DR WPI, 2001-442253/47.  
DR N-PSDE; AA159455.

PT novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -  
 XX  
 XX  
 PS  
 PS  
 Example 6; SEQ ID NO 3444; 10078pp; English.

The invention relates to human nucleic acids (AA157798-AA161369) and the encoded polypeptides (AA136642-AA142213) with neurotrophic, immunosuppressant and cyostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: immune system suppression, the activation/inhibition of activities such as: chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders.

Note: The sequence data for this patent did not form part of the printed specification.

Sequence 748 AA;

Query Match	Score	DB	Length
91.28;	2746;	22;	748;
Best Local Similarity	91.08;	22;	748;

[illegible]

```

QY      541 P-----RCHRPNSSTVARSWMSPVSAH---LSSTIPCT 573
          | | | | | | | | | | | | | | | | | | | |
Db      567 PPVPPTKQYLCPGLDAVLAIIRSPVFNHSLY--RTFYAM-tahgppitapvc 621

```

## RESULT

ID	AA	standard; Protein; 780 AA.
AA042085		

AC AAM42085;

DT 22-OCT-2001 (first entry)  
XX

Human polypeptide SEQ ID NO 7016.

KW Human; neotrophic; immunosuppressant; cytostatic; gene therapy; cancer  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;  
KW leukemia.

OS Homo sapiens.  
yy

PN WO200153312-A1  
XX

26-JUL-2001  
PD  
XX

26-DEC-2000; 2000WC-US34263.  
XX

PR 25-JAN-2000; 2000US-0488717  
PR 21-MAR-2000; 2000US-0552317  
PR 09-JUL-2000; 2000US-0598042  
PR 19-JUL-2000; 2000US-0620310  
PR 03-AUG-2000; 2000US-0653450  
PR 14-SEP-2000; 2000US-0662191  
PR 19-OCT-2000; 2000US-0693036  
PR 29-NOV-2000; 2000US-0727344

PA (HYSE-) HYSEQ INC.  
XX  
XX

PI Tang YT, Liu C, Asundi T, Chen R, Ma Y, Qian XB, Ren F, Wang D,  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J,  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries -

The invention relates to human nucleic acids (AA157798-AA161369) and the encoded polypeptides (AA386642-AA442213) with neurotrophic, immunosuppressant and cyostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders.

Note: The sequence data for this patent did not form part of the printed specification

Sequence 780 AA;







ins polypeptide comprises a novel human polymorphic 1 protein encoded by the hpl1 oncogene that is implicated in neoplastic disorders. Also claimed are: (1) an hpl1 polypeptide consisting of at least 22 contiguous amino acids; (2) an hpl1 fusion protein; antibodies which specifically bind to a hpl1 protein; (3) an expression construct for expressing all or a portion of hpl1 protein; (5) a homologously recombinant cell; (6) a method of



identifying neoplastic tissue of a human, where over-expression of the hpl gene identifies the tissue as being neoplastic; (7) a method to aid in the diagnosis or prognosis of neoplasia, where a difference in the hpl gene, mRNA, or protein between a first and second tissues indicates neoplasia in the first tissue; (8) a method to aid in detecting a genetic predisposition to neoplasia; (9) a method of identifying a human chromosome 12; and (10) a method of inducing a cell to de-differentiate by contact with a hpl gene or expression product. The products can also be used to inhibit hpl expression to suppress neoplasia, dysplasia, or hyperplastic cell growth. They can be used to treat e.g. tumours, amyloid hereditary ectodermal dysplasia, congenital alveolar dysplasia, epithelial dysplasia of the cervix, fibrous dysplasia of bone, or mammary dysplasia, endometrial, adrenal, breast, prostate, or thyroid hyperplasia or pseudo-epitheliomatous hyperplasia of the skin. Even in disorders in which hpl mutations are not implicated, down-regulation or inhibition of hpl expression can have therapeutic application. In these disorders, decreasing hpl expression can help to suppress tumours. Similarly, in tumours where hpl expression is not aberrant, effecting hpl down-regulation can suppress metastases.

Sequence 1004 AA;

Query Match	12.7%;	Score 382.5;	DB 19;	Length 1004;
Best Local Similarity	28.1%;	Pred. No. 1e-15;		
Matches 151;	Conservative 67;	Mismatches 218;	Indels 101;	Gaps 22

[illegible]

RESULT	7
AAW52830	
ID	AAW52830 standard; Protein: 1004 AA
XX	
AC	AAW52830;
XX	
DT	09-JUL-1998 (first entry)
XX	

A tumour suppressor gene called polyhomeotic 1 (*hph1*).

KW Tumour suppressor gene; polyhomeotic 1; hph1, human chromosome 12p13;  
KW identification; neoplastic tissue; cellular differentiation; diagnosis  
KW neoplasia; ss.

OS Homo sapiens.

PN W09807858-A1.

PD 26-FEB-1998.

PF 22-AUG-1997; 97WO-US14866

PR	06-FEB-1997;	97US-0036939
PR	03-MAR-1996	97US-0034340

PR 04-DEC-1996; 96US-0031569

PA (CHIR ) CHIRON CORP.

PI Randazzo F;

DR WPI; 1998-169162/15

XX  
XX

PT for diagnosis and th

XX  
XX

Claim 1; Pages 28-30; 40pp; English

CC The present sequence encodes a novel human tumour suppressor gene  
CC termed polyhemicin 1 (hph1). The hph1 gene maps to human chromosome  
CC 12p13, a region which is frequently lost in non-small cell lung cancer  
CC and breast cancer. A method of identifying neoplastic tissue of a  
CC human comprising comparing the expression of a hph1 gene in a tissue of  
CC a human suspected of being neoplastic with the expression of a hph1 gene  
CC in a tissue of the human which is normal. Under-expression of the hph1  
CC gene identifies the subject as having neoplastic tissue. The hph1  
CC oncogene functions to suppress neoplasia and dysplastic or hyperplastic  
CC cell growth as well as to induce cellular differentiation. The cDNA,  
CC protein and vectors can be used as diagnostic and therapeutic tools for  
CC proliferative and developmental disorders and to identify a p13 region  
CC of a human chromosome 12. They can be used for the detection, diagnosis  
CC or prognosis of neoplasia or for detecting a genetic predisposition to  
CC neoplasia. They can also be used to treat tumours.

Sequence	1004	AA
5Q		

Query Match	12.7%	Score 382.5;	DB 19;	Length 1004;
Best Local Similarity	28.1%;	Pred. No. 1e-15;		
Matches 151; Conservative	67;	Mismatches 218;	Indels 101;	Gaps 22

```
OY      60  OGTGGGPAAGATCMP-----PRGGTGLMGSGFGMAOQPMLSLQOPRGSTGMNHSM    115
        ||| | | | : | : | | | | | | | | | | | | | | | | | | | | | | | |
Db      260  gnlslsqagsgnslrpsmrggggqgnbhglqglrpsg--mggsgrkltgtvnpdrLa    317

OY      116  AVVSTATPTOTLQLOVALQOQOQOQOFQOQOQDALQOQOQOQOQOQFQAQGSAMQOQFO    175
        ||| | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      318  agtlvtvgsgsgaeasaakkaeadsgsgvnmllrtatrapsgtlllsatlutqlph    377

OY      176  AVVVQQQQQLQOQOQOQOQ--QHILKLYNQMOQOIQOQOQOQLRGLA-QILOQOQOQOQOQOQO    232
        :::::||||| : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      378  slldqqqghlqqkqvnlqqdalnh--qqgfqrnrqgllhtathldlaqqqqqqqqqq    435
        :::::||||| : | : | : | : | : | : | : | : | : | : | : | : | : |

OY      233  QOOQO-----ALEARPLQORMQORPORSQALRQOLOQMNNTONHQRORQORVAUQO    288
        ||| | | | : | | | | | | | | | | | | | | | | | | | | | | | |
Db      436  qqgqqqattltarprrvrpq--qvrpsqs-----qgqvdlvlvq-----rmlqss    480

OY      289  RSLQRLQSOTQGLVSQADALBQMILYTQTRFKGVARMV-----VOQRVOROYOQOQO    341
        ||| | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      481  pslrlprdaarkrpil-----lqskprvariklrpdlaqakmsaaqprr--rhlpvcy    529
```



QY 342 TAVCTAQAOMVAPVGVVSSSLPMLSSPSPGQOVOTPOSMPPPOPPPOPPSSQPN 401  
 Db 530 vltgptgagagajlajlaaav-----srgmptvq-sgqhlaspss 576  
 QY 402 NVSSGPAFSPSSFLPSPSPQSQ-----SPVTAATPQNFVSPSPGFLNTPVN 448  
 Db 577 qapgalqcppllaqgmflapvgtahvkgaltspvvaqvpafymqg---vhlpgk 633  
 QY 449 PSSVSPAGSSQAEEOOVLDKLTOLSKYIEPLRMINKIDKNEDRKDL-SKMSLIDL 507  
 Db 634 pqlt---avkrkadseerdvstclgsmtpakaspvaeapvmeksslgkaesvanv- 689  
 QY 508 TDPSCRCPLKTLQK-ETALEKTKNDKRCPLPFRPCHRPNSSTYASBSMPSMPTSA 564  
 Db 690 ---nanfpssel---val-----lpapsvp-----ppllamvstrgmgsdkpqa 727

## RESULT 8

AAB72673  
 ID AAB72673 standard; protein; 149 AA.

AC AAB72673;  
 XX

DT 09-MAY-2001 (first entry)  
 XX

DE Polyglutamine tract #2.  
 XX

KW Animal model: polyglutamine tract; neurodegenerative disorder: HDJ1;  
 KW heat shock protein 40; tetratricopeptide repeat protein 2; TPR2; stroke;  
 KW amyloid leukemia factor 1; MLF; human; fruit fly; Alzheimer's disease;  
 KW Parkinson's disease; CJD; BSE; Huntington's disease; head trauma; cancer.  
 XX  
 OS Drosophila sp.  
 XX  
 PN W0200112238-A1.  
 XX  
 PD 22-FEB-2001.  
 XX

PE 14-AUG-2000; 2000MO-0522496.  
 XX

PR 12-AUG-1999; 99US-0148933.  
 XX

PR 12-AUG-1999; 99US-0148934.  
 XX

PR 18-JAN-2000; 2000US-0177047.  
 XX

PR 19-MAY-2000; 2000US-0205720.  
 XX

(CALY) CALIFORNIA INST OF TECHNOLOGY.

PA Benzer S, Kazemi-Esfarjani P;  
 PI  
 XX

DR WPI; 2001-147537/15  
 XX

PT Identifying genes of other compounds that modulate polyglutamine  
 PT toxicity, useful for treating Alzheimer's disease, Parkinson's disease  
 PT and Creutzfeldt-Jakob disease -  
 XX  
 PS  
 XX

Example 2; Fig 1B; 775pp; English.

XX The present invention describes a method of screening for genes which  
 CC modulate polyglutamine toxicity using animal models with polyglutamine  
 CC sequences that cause toxicity in the animal. The model is preferably  
 CC Drosophila, and toxic polyglutamine sequences include the human and  
 CC Drosophila heat shock protein 40/HDJ1, tetratricopeptide repeat protein 2  
 CC (TPR2) and myeloid leukemia factor 1 (MLF) genes. The model is useful  
 CC for identifying treatments for neurodegenerative and proliferative  
 CC disorders, including Alzheimer's disease, Parkinson's disease,  
 CC Creutzfeldt-Jacob's disease (CJD), bovine spongiform encephalopathy  
 CC (BSE), Huntington's disease, Machado-Joseph disease, spinocerebellar  
 CC ataxias, dentatorubropallidoluysian atrophy, Kennedy's disease, stroke,  
 CC head trauma and cancer.  
 XX  
 SQ Sequence 149 AA;

Query Match 12.5%; Score 376.5; DB 22; Length 149;  
 Best Local Similarity 55.8%; Pred. No. 3,1e-16;  
 Matches 91; Conservative 5; Mismatches 46; Indels 21; Gaps 2;

QY 99 LSGQPPPGTSGMFAHMAVSTATPOTOLQLOVALQOQOQOQOQOQOQOQOQOQ 158  
 Db 1 mgsqpp-----stpqgqgqgqgqgqgqgqgqgqgqgqgqgqgqgqgqgq 44

QY 159 QQQQFOAQQSAAQQFOAQQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ 218  
 Db 45 qqqqgq 99

QY 219 QLOQ 261  
 Db 100 qqqqgq 142

## RESULT 9

AAB69612  
 ID AAB69612 standard; protein; 155 AA.

AC AAB69612;  
 XX

DT 30-APR-2001 (first entry)  
 XX

DE Huntingtin accumulation inhibitor peptide HD-Q104-Myc-HIS6.  
 XX

KW Neurological disorder; Huntington's disease; Alzheimer's disease;  
 KW Parkinson's disease; prion disease; frontotemporal dementia;  
 KW amyotrophic lateral sclerosis; spinal and bulbar muscular atrophy;  
 KW dentatorubral-pallidoluysian atrophy; spinocerebellar ataxia type 1;  
 KW SCA3; SCA4; SCA5; SCA6; SCA7; protein accumulation; intrabody.  
 XX  
 OS Synthetic.  
 XX  
 PN W0200106989-A2.  
 XX  
 PD 01-FEB-2001.  
 XX

PE 24-JUL-2000; 2000MO-US20131.  
 XX

PR 27-JUL-1999; 99US-0146047.  
 XX

PR 21-JUL-2000; 2000US-0620955.  
 XX

PA (HUST/) HUSTON J S.  
 PA (MESS/) MESSER A.  
 PA (LECE/) LECERF J.  
 XX  
 PI  
 XX

PI Huston JS, Messer A, Lecerf J;  
 XX

DR WPI; 2001-182700/18.  
 XX

PT Inhibiting intracellular polypeptide accumulation, useful for treating  
 PT neurological disorders, e.g. Alzheimer's disease, comprises contacting  
 PT the polypeptide with a specific intrabody -  
 XX  
 PS  
 XX

Disclosure: Page 99; 108pp; English.

XX The present invention describes a method for inhibiting the formation of  
 CC aggregates of certain proteins, involving contacting the protein with a  
 CC binding molecule known as an intrabody. Proteins to be bound include  
 CC those associated with neurological disorders, and so the method can be  
 CC used in the prevention of diseases such as Alzheimer's, Parkinson's and  
 CC Huntington's diseases, prion disease, frontotemporal dementia,  
 CC amyotrophic lateral sclerosis, spinal and bulbar muscular atrophy,  
 CC dentatorubral-pallidoluysian atrophy, spinocerebellar ataxia type 1  
 CC (SCA1), SCA2, SCA3, SCA4, SCA5, SCA6 and SCA7.  
 XX  
 SQ Sequence 155 AA;

Query Match

11.6%; Score 350.5; DB 22; Length 155;



Query Match:	10.9%	Score 328.5;	DB 22;	Length 121;
Best Local Similarity	69.4%;	Pred. No. 2, 2e-13;		
Matches	77;	Conservative	1;	Mismatches 26;
			Indels	7;
			Gaps	1















'Fri Mar 1 09:12:01 2002

us-09-668-119-3.rag

---



—  
—







```

OY 116 AVSTATPOTOTLOLOVNAALOOOOOOOPOPOOOOAAALOOOOOOOPOPOOAAOQSMQOPO 175
Db 318 AOTVTVSGSOTEAISAARKAEADGSGQVWGNMILRTAPRSQTLISATYTQORH 3777
OY 176 AVVOOOOOLOOOOOPOO--ONIKLNHMOOQIOOOOOOQRIA-QLOOOOOOOOQOQ 2322
Db 378 SLIOOOOQIHLQOKVUVIQQOIALHN--QOQFORQOGLHTAHNLDAOOOOOQOQ 4355
OY 233 OOOO--ALBAOPHOORPMQOORPSSALPOOLOMHNHTONORRPOORPVAQO 288
Db 436 OOOOQOATTLTAPOQVPRPO--QVPRSQS-----QOAOVLVQ-----PMLQSS 480
OY 289 PSOLPQSOQOPLVQOALQOGMLYTOPRLKFRAMV-----VOQRPVQVOOQ 3411
Db 481 PLSLPRDAARPRIT-----IOSKPPAPILKRPQLGAKMSAODRP--PIHIVQV 529
OY 342 TAVOQAQAQVAVRVOVQSGSLPMLSSPBGQOVOTPOSMRPPROPPOGPOSSOPNS 401
Db 530 VGTROPQTAQOALCLQLAOALAAR--TSKQMPGTVO--SGOHLASSPSS 576
OY 402 NVSSGPAPSSPFLSPSPQO-----SPVARTPQNPVSPSPGLNTPVN 448
Db 577 QAPGALOCPPTLAIGMTLAPVGTAVVKGAGTTSSPVAQVAPAKWOS--VHLPGK 6333
OY 449 PSSVSPAGSSQAEHQOYLKLOLSKYIEPLRMINKIRKNEDRKDI-SKAKSLDIL 5077
Db 634 PQLT--AVKRKADSEERDQVSTLGSMLPAKSPVAESPKVDEKSLERKEASVAN-- 669
OY 508 TDPKRCPLKLOKELALEKLKNDMCRPLRPRRCRPNSSITYASRWPMSPTSA 564
Db 690 ---NANPSSSEL---VAL-----TPAVSV---PTLAMVSNQMDSDSPQA 727

RESULT 2
US-08-991-300-2
Sequence 2, Application US/08991300
Patent No. 5973225
GENERAL INFORMATION:
APPLICANT: D'OVIDIO, RENATO
APPLICANT: PORCEDDU, ENRICO
APPLICANT: MERCHITELLI, CINZIA
APPLICANT: CARDELLI, LUISA EROLDI
TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF A GENE
NUMBER OF INVENTION: ENCODING A LOW MOLECULAR WEIGHT GLUTENIN
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSER: P.O.C., SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/991,300
FILING DATE: 16-DEC-1997
CLASSIFICATION: 830
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IT MI 96/A 002663
FILING DATE: 19-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: OSLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 2264-0201-0X
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-1220
INFORMATION FOR SEQ ID NO: 2:

```

```

? SEQUENCE CHARACTERISTICS:
? LENGTH: 369 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: protein
US-08-991-300-2

Query Match          10.3% Score 311.5; DB: 2; Length 369;
Best Local Similarity 32.1%; Pred No. 5,6e+15;
Matches 121; Conservative 32; Mismatches 149; Indels 75; Gaps 16

QY 134 LQQNQDQDQDQDQDQDQALDQDQDQ--QDQDQDQDQDQDQAFAYVQDQDQ--LQDQD 188
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   6 LEKSSQDQLPRPQDTLSHHHQDLPDQDHPDQDPCSSQDQDQDPRLSQDQDPPFSQDQ 65
QY 189 ---QQQNHKLKHND-NQQT-----QQDQDQLDRDLQLDQDQDQDQDQ---QDQDQ 236
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   66 PPSQDQDQVLPLPDQPSPFSQDQLPPSQQDQDPPFSQDQDQVLPLPDQPSPFSQDQL 125
QY 237 ALEADPTI--QDPPMQDPPR--SQLPLDQLQMHNTHQDHPDQDPPRYAQDNQSDLP 293
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   126 PFSQDQVLPLPDQDPPFSQDQDPPFSQDQLPPFSQD-----QDPLPDQDPPFSQDQDPI 178
QY 294 PQSGTQPLVSQAQDALPGMILYTQPLKFKYRAMVVQDPPVQDPPVQDGTAVQTAAQAM 353
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   179 PQ---QPPFSQDQ---QPVLLDQDQIPFVNHPILLQDLNPKCKVFLLQDCSPWAMPQS---L 228
QY 354 APGVQVSSSLPMLSSPSPGQDVQTPQ-----SMPPRPSS 389
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   229 ARSQMLQSSSCHVMQDQCQQLQDLPIDQDSRYEALRALIVYSITLDQEQQVGSIQTQDQDP 288
QY 390 POPGPSPSP--SNVSSGPAPSPSPSPSPSPSQ-----SPTATTPGNFVPSP 440
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   289 QQLQGVCVSPQDQDQQLDQDQDQLAHGFTLDPHQDQLAQLEVNTSLALT-----LPTM 343
QY 441 GLPLTPPNVSPSVMPAG 457
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   344 CNMNVPLRYTRTPRG 360
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 3
US-08-918-914-4
? Sequence 4, Application US//08918914
? Patent No. 5876963
? GENERAL INFORMATION:
? APPLICANT: Mitchell, Peter
? APPLICANT: Hutchinson, Nancy
? APPLICANT: Lawton, Michael
? APPLICANT: Magna, Holly
? APPLICANT: Yocum, Sue
? TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHORYLASE
? NUMBER OF SEQUENCES: 4
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Incyte Pharmaceuticals, Inc.
? STREET: 3174 Porter Dr.
? CITY: Palo Alto
? STATE: CA
? COUNTRY: USA
? ZIP: 94304
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: DOS
? SOFTWARE: FASTSEQ for Windows Version 2.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US//08/918,914
? FILING DATE: Filed Herewith
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE:
```







Page 4

Query Match 9.6%; Score 288; DB 2; Length 1162;  
Best Local Similarity 35.9%; Pred. No. 1.1e-12;  
Matches 83; Conservative 28; Mismatches 90; Gaps 7;

RESULT 6  
US-09-625-188-20

Query Match	9.58;	Score 286;	DB 4;	Length 729;
Best Local Similarity	28.08;	Pred. No. 8.4e-13;		

RESULT 7  
US-08-194-468-2  
; Sequence 2, Application US/08194468

```

1  GENERAL INFORMATION:
2  APPLICANT: Montminy, Marc R.
3  TITLE OF INVENTION: ASSAYS FOR THE IDENTIFICATION OF
4  TITLE OF INVENTION: COMPOUNDS WHICH INHIBIT ACTIVATION OF CAMP AND MITOGENE
5  TITLE OF INVENTION: RESPONSIVE GENES
6  NUMBER OF SEQUENCES: 3
7  CORRESPONDENCE ADDRESS:
8  ADDRESSEE: Pretty, Schroeder, Buessgemann & Clark
9  STREET: 444 South Flower Street, Suite 2000
10 CITY: Los Angeles
11 STATE: California
12 COUNTRY: USA
13 ZIP: 90071
14
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE: Floppy disk
17 COMPUTER: IBM PC compatible
18 OPERATING SYSTEM: PC-DOS/MS-DOS
19 SOFTWARE: PatentIn Release #1.0, Version #1.25
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER: US/08/194,468
22 FILING DATE: 10-FEB-1994
23 CLASSIFICATION: 435
24 ATTORNEY/AGENT INFORMATION:
25 NAME: Reiter, Stephen E.
26 REGISTRATION NUMBER: 31,192
27 REFERENCE/DOCKET NUMBER: P41 9672
28 TELECOMMUNICATION INFORMATION:
29 TELEPHONE: (619)-546-4737
30 TELEFAX: (619)-546-9392
31 INFORMATION FOR SEQ. ID NO.: 2:
32 SEQUENCE CHARACTERISTICS:
33 LENGTH: 2441 amino acids
34 TYPE: amino acid
35 TOPOLOGY: linear
36 MOLECULE TYPE: protein
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966

```



Query Match 9.4%; Score 283; DB 1; Length 2441;  
 Best Local Similarity 27.0%; Pred. No. 5.8e-12;  
 Matches 141; Conservative 40; Mismatches 165; Indels 176; Gaps 22;

```

QY 74 GMPRPGQSLGMSGFMAGMOPMSLSGPPGTSGMAFPHMAVSTATPOTQLOQOVA 133
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1981 GMP-----GRDGMGTSPSSQMTVGLNVRPNVSG-----PVMSMPG---QWQAP 2026
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 134 LQOQO-----QOQOFOQOQOQAL-----QOQOQ 156
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2027 IPQOQPMGMPRVMSMQDAVAAGPRMNVOPNRSTSPSALQDLRTLKSPSSPQOQOQ 2086
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 157 -----QOQOQOFOQOQSAMQOQFOAVVQOQOQLOQOQOQOHLTKL--- 197
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2087 VLNLKSNPOLMAAFIKORTAKYVANQPGMQPQGLQSQGMQPGMGHQPSLQNLNAM 2146
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 198 -----LPIQSSMGMAAPMGQLGQM---GQPLGADSTPN---IQALQQRILQOQOQMKQ 2252
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2147 QAGVRPVPVPPOPRAMGSLNPGQALNINPGRNMTNMPDYREKVRQLDHOQOQOQ 2206
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 226 QOQOQOQOQOQOQALF---AOPRIQOPRMOORPPPSQALPQOLOMHNTHOPRPPQOPRV 284
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2207 QOQOQOQOQOQOQNSASLAGMGHSGFOQPGQPGYARMAQOQRM---QOH----- 2252
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 285 AONPSQLPPOSQOTPLVSAQALPGQMLYTOPRLKFRARVYVQOPVQVQOQOQTA 344
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2253 -----LPIQSSMGMAAPMGQLGQM---GQPLGADSTPN---IQALQQRILQOQOQMKQ 2302
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 345 QTAQAQ-----MVA-----PGVQVSSSLMILSPSPGQOQVOTPOSMPRPQ 387
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2303 QIGSPGQPNMPSPOQHMLSGOPQASHLPQOQIATSLSNQVSPAPVQSPR---POSQPPHSS 2361
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 388 PSP--QPGSPSQPNMPSVSGPAPSPSSFLPSPSPQ-----PSQSPYARTPQNFVSVP 440
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2362 PSPRIQF-QPSPHNVSTQGTPIHGLAVTMASMDQGHGNGPDS---AMLPQ----- 2410
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 441 GPLNTPVNP--SSVMSFAGSSQAEQOYLKIKOLSKYIEPL 480
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2411 --LNTPNRSALSSLSLIVGDTGD-----TLEKFVEGL 2441
    : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 8  
 US-08-961-739-2  
 ; Sequence 2, Application US/08961739A  
 ; Patent No. 6063583  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Montminy, Marc R.  
 ; TITLE OF INVENTION: Methods for Treating Diabetes Mellitus  
 ; FILE REFERENCE: SAKI650-1  
 ; CURRENT APPLICATION NUMBER: US/08/961,739A  
 ; CURRENT FILING DATE: 1997-10-31  
 ; EARLIER APPLICATION NUMBER: US 194,468  
 ; EARLIER FILING DATE: 1994-02-10  
 ; NUMBER OF SEQ ID NOS: 4  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 2  
 ; LENGTH: 2441  
 ; TYPE: PRT  
 ; ORGANISM: Mus  
 ; FEATURE:  
 ; NAME/KEY: VARIANT  
 ; LOCATION: (1)...(2441)  
 ; OTHER INFORMATION: Xaa = Any Amino Acid  
 US-08-961-739-2

Query Match 9.4%; Score 283; DB 3; Length 2441;  
 Best Local Similarity 27.0%; Pred. No. 5.8e-12;  
 Matches 141; Conservative 40; Mismatches 165; Indels 176; Gaps 22;

```

QY 134 LQOQO-----QOQOFOQOQOQAL-----QOQOQ 156
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2027 IPQOQPMGMPRVMSMQDAVAAGPRMNVOPNRSTSPSALQDLRTLKSPSSPQOQOQ 2086
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 157 -----QOQOQOFOQOQSAMQOQFOAVVQOQOQLOQOQOQOHLTKL--- 197
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2087 VLNLKSNPOLMAAFIKORTAKYVANQPGMQPQGLQSQGMQPGMGHQPSLQNLNAM 2146
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 198 -----LPIQSSMGMAAPMGQLGQM---GQPLGADSTPN---IQALQQRILQOQOQMKQ 2252
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2147 QAGVRPVPVPPOPRAMGSLNPGQALNINPGRNMTNMPDYREKVRQLDHOQOQOQ 2206
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 226 QOQOQOQOQOQOQALF---AOPRIQOPRMOORPPPSQALPQOLOMHNTHOPRPPQOPRV 284
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2207 QOQOQOQOQOQOQNSASLAGMGHSGFOQPGQPGYARMAQOQRM---QOH----- 2252
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 285 AONPSQLPPOSQOTPLVSAQALPGQMLYTOPRLKFRARVYVQOPVQVQOQOQTA 344
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2253 -----LPIQSSMGMAAPMGQLGQM---GQPLGADSTPN---IQALQQRILQOQOQMKQ 2302
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 345 QTAQAQ-----MVA-----PGVQVSSSLMILSPSPGQOQVOTPOSMPRPQ 387
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2303 QIGSPGQPNMPSPOQHMLSGOPQASHLPQOQIATSLSNQVSPAPVQSPR---POSQPPHSS 2361
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 388 PSP--QPGSPSQPNMPSVSGPAPSPSSFLPSPSPQ-----PSQSPYARTPQNFVSVP 440
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2362 PSPRIQF-QPSPHNVSTQGTPIHGLAVTMASMDQGHGNGPDS---AMLPQ----- 2410
    : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 441 GPLNTPVNP--SSVMSFAGSSQAEQOYLKIKOLSKYIEPL 480
    ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2411 --LNTPNRSALSSLSLIVGDTGD-----TLEKFVEGL 2441
    : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 9  
 US-08-227-536-2  
 ; Sequence 2, Application US/08227536  
 ; Patent No. 5658784  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Eckner, Richard  
 ; APPLICANT: Ewen, Mark  
 ; TITLE OF INVENTION: NUCLEIC ACID, ENCODING TRANSCRIPTION  
 ; TITLE OF INVENTION: FACTOR P300 AND USES OF P300  
 ; NUMBER OF SEQUENCES: 13  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes  
 ; STREET: Ten Post Office Square  
 ; CITY: Boston  
 ; STATE: MA  
 ; COUNTRY: US  
 ; ZIP: 02109  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/227,536  
 ; FILING DATE: 14-APR-1994  
 ; CLASSIFICATION: 436  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Williams Ph.D., Kathleen A.  
 ; REGISTRATION NUMBER: 34,380  
 ; REFERENCE/DOCKET NUMBER: DFCI-308XX  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (617) 542-2290  
 ; TELEFAX: (617) 451-0313  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2414 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear



















Fri Mar 1 09:12:01 2002

us-09-668-119-3.ra1

---

Page 10







```
alignment_scores:
```

Quality:	2988.00	Length:	580
Ratio:	5.161	Gaps:	1
Percent Similarity:	99.828	Percent Identity:	99.483

```
alignment_block:
file=00-668-110-2-11100111
```

Align seg 1/1 to: AAI59455 from: 1 to: 2556

[illegible]

1088	CAGCCACAGAGCTCCGGCCCTCCAGAGCTCGCCCAAGACAGTACACAGAT	1137
267	TAASHASTThGlnHisGlnProProProGlnProGlnGlnProProy	284
1138	GCATCACACACAGCACCACAGCCGGACACACAGCCCAAGAGCTCCAG	1187
284	ALAAGlnAsnGlnProSerGlnLeuProGlnInSerGlnThGlnPro	300
1188	TTGCTCAGAACCAACCATACACACTCCCGCAAGTCCCAAGCCAGCT	1237
301	LeuValSerGlnAlaGlnAlaLeuProGlyGlnMetLeuYrThGlnPr	317
1238	TTGGTGTCAAGAGCGGAGACCTCTCCCTGGACAAAGTTGTATTCCCAAC	1287
317	oProLeuYrSphenValArgAlaProMetValAlcGlnInbProProValG	334
1288	ACCACATGAATTTGTCGAGGTCGATGTTGGTCAGAGCCCAAGCTGC	1337
334	1nProGlnValGlnGlnGlnInbTrpAlaValGlnThAlaGlnAlaA	350
1338	AGCCACAGTGTGACAGACAGCAGACAGATACACAGCTACAGCGTCC	1387
351	GlnMetValAlaProGlyValGlnValSerGlnSerLeuProMetLe	367
1388	CAGATGTGGTCCCGGAGTCCAGGTTCAGCCAGACAGACCTCCCATGCT	1437
367	uSerSerProSerProGlyGlnGlnValGlnThbProGlnSerMetbPro	384
1438	GHCTGTGGCGGTACAGGGGCGACAGAGTGCACACCCCGAGTCGATGGCC	1487
384	roProProGlnProSerProGlnProGlyGlnProSerSerGlnProAsn	400
1488	CHCCTCCCGAGCGGTCCCGCAGCCCGGCGACGCCACTCACAACCCAC	1537
401	SerAsnValSerSerGlyProAlaProSerProSerSerPheLeuProSe	417
1538	TCCAAACGTACGCTCGGCCCTCGCCCATCTCCCAAGTACTTCCTGGCCAG	1587
417	ProSerProGlnInProSerGlnSerProValThbAlaArgThbProGlnA	434
1588	CCCCCTACCGAGCCCTCCCAAGCCCAAGTACGGCGCGGACCCCAACA	1637
434	snPheSerValProSerProGlyProLeuAsnThbProValAsnProSer	450
1638	ACTTCAAGTGTCCCTCCATCCTGGACCTTTAAACACACCTGTGAACCCGAGC	1687
451	SerValMetSerProAlaGlySerSerGlnAlaGlnLeuGlnGlnYrLe	467
1688	TCTGTTCATGAGCCCAAGCTGGGTCCAGCCAGGCTGTGAGGACGACGATACCT	1737
467	uAspYrLeuLeuGlnLeuSerYrYrLeGlnProLeuAlaArgMet	484
1738	GGACACAGCTGAAAGCAGCTGTGGAAGTCAATCGAGCCCTCGCGCGCATGA	1787
484	LeuAsnYrLeaAspLysAsnGlnAspArgYrGlyLysAspLeuSerYrMet	500
1788	TCAACAAAGATGACACAGACCAACGAAAAAAGAGACCTGAGTAAGATG	1837
501	LysSerLeuLeuAspPheLeuThbAspProSerYrAspGlyProLeuYr	517
1838	AAGAGCCTTGTGGACATTGTGACAGACCCCTCGAAGGCGTGTCCCTCGAA	1887
517	sThrLeuGlnLeuGlyGlnLeuAlaLeuGlnYrLeuLysAsnAspMet	533
1888	GACCTTGGAAAAAGTGTAGATCGCCCTGGAGAAACTCAAAATGACATGG	1937
534	ArgGlyProLeuProHisArgProAlaGlyGlnHisArgProAsnSerTh	550
1938	CGGAGCCCACTCCCAACCGCCCGGTGGCCACGACCAACACGCGATAC	1987
550	TTrpAlaSerArgSerTrpMetProSerTrpProThbSerAlaHisLeu	567



1988 CTATGCCAGCCGCTCCTGATGCGCTGCGCCACATCCGCTCACCCTGT 2037

567 eSrThrIleProGlyThrAlaHisSerPheGlnPro 579

2038 CTTCACACATTCCTCCGTACCGCACATTCGTTCCAGCCA 2075

seq\_name: /SID57/gcgdata/geneseq/geneseqn/MA2001.DAT:AA161241

seq\_documentation\_block:

ID AA161241 standard; CDNA; 3438 BP.

AA161241;

22-OCT-2001 (first entry)

Human polynucleotide SEQ ID NO 5230.

Human; nocrotic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokine; thrombolytic; drug screening; arthritis; inflammation; leukaemia; ss.

Homo sapiens.

W020015312-A1.

26-JUL-2001.

26-DEC-2000; 2000WO-US34263.

21-JAN-2000; 2000US-0488725.

25-APR-2000; 2000US-0552317.

09-JUL-2000; 2000US-0598042.

19-JUL-2000; 2000US-0620312.

03-AUG-2000; 2000US-0653450.

14-SEP-2000; 2000US-0662191.

19-OCT-2000; 2000US-0693036.

29-NOV-2000; 2000US-0727344.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;

Zhao QH, Zhou P, Goodrich R, Drmanac RT;

WPI; 2001-442253/47.

P-PSDB; AAM42085.

Novel nucleic acids and polypeptides, useful for treating disorders

such as central nervous system injuries -

Claim 1; SEQ ID NO 5230; 10078bp; English.

The invention relates to human nucleic acids (AA157798-AA161369) and

the encoded polypeptides (AAM38642-AA42213) with nocrotic,

immunosuppressant and cytostatic activity. The polynucleotides are useful

in gene therapy. A composition containing a polypeptide or polynucleotide

of the invention may be used to treat diseases of the peripheral nervous

system, such as peripheral nervous injuries, peripheral neuropathy and

localised neuropathies and central nervous system diseases, such as

Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

utilization of the activities such as: Immune system suppression,

Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

and thrombolytic activity, cancer diagnosis and therapy, drug screening,

assays for receptor activity, arthritis and inflammation, leukaemias and

C.N.S disorders.

Note: The sequence data for this patent did not form part of the printed

specification.

Sequence 3438 BP; 771 A; 1143 C; 932 G; 592 T; 0 other;

alignment\_scores:

Quality: 2886.00 Length: 587

Ratio: 4.993 Gaps: 8

Percent Similarity: 98.467 Percent Identity: 97.956

alignment\_block:

US-09-668-119-3 x AA161241

Align seg 1/1 to: AA161241 from: 1 to: 3438

1 MetArgLysAlaGlyValAlaHisSerLysSerLysAspMetGlu 17

338 ATGAGAAAGCTGTGTGGCACAGTAATCCAGCAAGATATGAGAG 387

17 HisValPheLeuLysAlaLysThrArgAspLysLeuSerLeuVal 34

388 CCATGTTTCCGAAAGCCAGACCCGGGCAATACCTTCTCTGTGG 437

34 LArgLeuIleIleHisPheArgAspIleHisAsnLysSerGlnAla 50

438 CCAAGCTCATTTCCATTTTCAGACATTCATACAGAAATCTCAGCT 487

51 SerValSerAspProMetAsnAlaLeuGlnSerLeuThrGlyProAl 67

488 TCCGTCAGTATCTTATGATGACACTCCAGAGCTGAGTGGGACCTGC 537

67 aaAGlyAlaAlaGlyIleGlyMetProProArgGlyProGlyGlnSerL 84

538 TGGGGAGCGCGCTGGAAATGGCATGCTCTCGGGGCCCGGACAGTCTC 587

84 euAGlyLysMetGlySerPheGlyAlaMetGlyGlnProMetSerLeuSer 100

588 TGGGGCGGATGGGTAGCTTGTGGCATGGAGCAGCAATGCTCTCTCA 637

101 GlyGlnProProGlyThrSerGlyMetAlaProHisSerMetAlaVal 117

638 GGGCAGCCGCTCTCTGGGACCTCGGGGATGGCCCTCAGCAGATGGCTCT 687

117 ValSerThrAlaThrProGlnThrGlnLeuGlnLeuGlnValAla 133

688 CGTGTCTACGGCACTCCACAGCCAGCTGCAGCTCCAGAGTGGCGG 737

134 LeuGlnGlnGlnGlnGlnGlnGlnGlnPheGlnGlnGlnAla 150

738 CTGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCG 787

150 LaleuGlnGlnGlnGlnGlnGlnGlnGlnGlnPheGlnAlaG 166

788 CGCTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCTCA 837

166 ngInSerAlaMetGlnGlnGlnPheGlnAlaValValGlnGlnGln 182

838 GCAAGTGCATGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 887

182 ngInLeuGlnGlnGlnGlnGlnGlnGlnGlnHisLeuLeuHis 198

888 GCAGCTTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCT 937

199 HisGlnAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnArg 215

938 CATCAAAATCAGCAACAGATACAGCAGCAGCAGCAGCAGCAGTGGAGAT 987

215 eaAGlnLeuGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 232

988 AGCAGCAGTGCAGTCCACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1037

232 lngGlnGlnGlnGlnAlaLeuGlnAlaGlnProProIleGlnGlnProPro 248

1038 AGCAGCAGCAGCAGCAGCTTTCAGGCCCCAGCCAGCAATTCAGCAGCAGC 1087

249 MetGlnGlnProGlnProProSerGlnAlaLeuProGlnGlnGln 265



```

1088 ATGCAGCAGCCACAGTCTCCGCCCCCTCCAGGCTCTGCCCCAGCAGCTGCA 1137
265 NGIMETHSHSHSTHrGlnHSHSHSGLnProProProGlnProGlnGlnP 282
1138 GCAGATGATCAGACACACAGCAGCCAGCCAGCAGCAGCCAGCAGCAGC 1187
292 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 298
1188 CTCCAGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1237
299 GlnProLeuValSer3lnAlaGlnAlaLeuProGlyGlnMetLeuTyrTh 315
1238 CAGCCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1287
315 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1337
1288 CCAACACACACTGAAATTTGTCCAGGCTCCGATGTGTGTGTGTGTGTGT 1337
332 rovalGlnProGlnValGlnGlnGlnGlnHrAlaValGlnThrAlaGln 348
1338 CAGTGCAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1387
349 AlAlaGlnMetValAlaProGlyValGlnValSerGlnSerLeuPr 365
1388 GCTGCCAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1437
365 oMetLeuSerSerPrSerProGlyGlnGlnValGlnThrProGlnSerm 382
1438 CATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1487
382 etProProProProGlnProSerProGlnProGlyGlnProSerSerGln 398
1488 TGCCCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCAG 1537
399 ProAnSerAnValSerSerGlyProAlaProSerProSerSerPheLe 415
1538 CCCAGCTCCCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1587
415 upProSerProSerPrGlnProSerGlnSerProValThrAlaArgThr 431
1588 GCCCAGCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 1637
432 ProGlnAnSerPrSerValProSerProGlyProLeuAnThrProValAs 448
1638 CCAGAGAACTTCACTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCGA 1687
448 nProSerSerValMetSerProAlaGlySerSerGlnAlaGlnGlnGln 465
1688 CCCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1737
465 InTyrLeuAspLysIleuLysGlnLeuSerLysTyrIleGlnProLeuArg 481
1738 AGTACCTGGCAAGCTCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 1787
482 ArgMetIleuLysIleuLysAspLysAspArgLysLysLysLysLys 498
1788 CGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1837
498 rLysMetLysSerLeuLeuAspLysLeuThrAspProSerLysArgCys 515
1838 TAAGATGAAGAACCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1887
515 rLysMetLysThrLeuGlnLysCysGlnAlaLeuGlnLysLysLysLys 531
1888 CCTGTGAAGACCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1937
532 AspMetArgCysProLeuProHisArgProArgCysHisArgProAsn 548
1938 GACATGGCGGTGCCCATTCCTCCAGCCGCTCCCTCCCTCCCTCCCTCC 1987
548 etSerThrTyrAlaSerArgSerTrpMetProSerTrpProThrSerAla 564

```

```

1988 GCAGTACTATGCCAGCCGCTCTCGATGCCCTCTGTCGCAACATCCGCT 2037
565 HisLeuSerSerThrIleProCysThrAlaHisSerPheGlnPro 579
2038 CACCTGTCTTCAACATTCCTCTACCGCAGCATCTCTCCAGGCA 2082
seq_name: /SIDS2/gcdata/geneseq/geneseq/NA2000.DAT: AAC59039
seq_documentation_block:
ID AAC59039 standard; cDNA; 1388 BP.
XX
AC AAC59039;
XX
DT 02-FEB-2001 (first entry)
XX
DE Human secreted protein coding sequence SEQ ID NO: 60.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antinflammatory; antitumor;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; ss.
XX
OS Homo sapiens.
XX
PN WO200055175-A1.
XX
PD 21-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-US06049.
XX
PR 12-MAR-1999; 99US-0124144.
PR 11-JUN-1999; 99US-0138574.
PR 03-DEC-1999; 99US-0168667.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI: 2000-638175/61.
XX
DX P-PSDB: AAB27609.
XX
PM Isolated nucleic acid molecule encoding a human secreted protein is
XX used in preventing, treating or ameliorating a medical condition.
XX
PS Claim 1; Page 372; 428pp; English.
XX
CC The invention relates to the isolation of genes AA58990-A59039 encoding
CC 50 human secreted proteins AAB27560-B27609. The genes can be used to
CC generate fusion proteins by linking to the gene for the human
CC immunoglobulin G Fc portion (SEQID) for increasing the stability of
CC the fusion protein as compared to the human protein only. The genes and
CC proteins are useful for preventing, ameliorating or treating medical
CC conditions, e.g. by protein or gene therapy. The genes are isolated
CC from a range of human tissues disclosed in the specification. The
CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune;
CC hemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d)
CC wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
XX
SQ Sequence 1388 BP; 364 A; 378 C; 343 G; 297 T; 6 other;

```

alignment\_scores: Quality: 1597.00 Length: 319  
Ratio: 5.006 Gaps: 0



Percent Similarity: 100.000 Percent Identity: 98.746

alignment\_block:

US-09-668-119-3 x AAC59039

Align seg 1/1 to: AAC59039 from: 1 to: 1388

```

1 MetArgLysAlaGlyValAlaHisSerLysSerLysAspMetLys 17
  |||
99 ATAGGAAAGCTGTGTGGACA.AGTAAATCCAGCAGAGATATGGAGAG 147
17 HisValPheLeuLysAlaLysThrArgAspGluTyrLeuSerLeuValA 34
  |||
148 CCATGTTTCTCTGAGGCCAGCAGCCGGGAGCATACCTTCTCTGCTGG 197
34 LaArgLeuIleIleHisPheArgAspIleHisAsnLysSerGlnAla 50
  |||
198 CCAGGCTCATTCATTCATTCGAGCATTCATTAACAAGAAATCTCAAGCT 247
51 SerValSerAspProMetAsnAlaLeuGlnSerLeuThrGlyGlyProAl 67
  |||
248 TCCGTAGATATCTATGAATGCATCCAGAGCTGATGCGGAGCTGCG 297
67 aAlaGlyAlaAlaGlyIleGlyMetProProArgGlyProGlyGlnSerL 84
  |||
298 TCGGGAGAGCGCGTGAATTGGCATGCTCTCGGGGCGCGGAGACATCTC 347
84 euGlyGlyMetGlySerPheGlyAlaMetGlyLysProMetSerLeuSer 100
  |||
348 TGGGCGGAGTGGTACGCTTGTGTCATGGAGACGCCAATGCTCTCTCA 397
100 GlyGlnProProGlyThrSerGlyMetAlaProHisSerMetAlaVal 117
  |||
398 GGGGAGCGGCTCTGGGACCTCGGGAGTGGCCCTCAGCAGATGGGTGT 447
117 ValSerThrAlaThrProGlnThrGlnLeuGlnLeuGlnAlaValAla 134
  |||
448 CGTGTCTACGGCAACTCCACAGACCCAGCTGAGCTCAGCAGGTGGCGC 497
134 euGlnGlnGlnGlnGlnGlnGlnPheGlnGlnGlnGlnGlnAla 150
  |||
498 TGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 547
151 LeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 167
  |||
548 CTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 597
167 nSerAlaMetGlnGlnGlnPheGlnAlaValGlnGlnGlnGlnGln 184
  |||
598 GAGTGCATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 647
184 euGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 200
  |||
648 TCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 697
201 AsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 217
  |||
698 AATAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 747
217 nLeuGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 234
  |||
748 GCTGAGCTTCAACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 797
234 GlnGlnAlaLeuGlnAlaGlnProProIleGlnGlnGlnGlnGlnGln 250
  |||
798 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 847
251 GlnProGlnProProSerGlnAlaLeuProGlnGlnGlnGlnGlnGln 267
  |||
848 CAGCCACAGCCTCCGCTCCAGGCTCTGCCAGCAGCAGCAGCAGCAG 897
267 ThrHisThrGlnHisGlnProProGlnGlnGlnGlnGlnGlnGln 284
  |||

```

```

898 GCATCCACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 947
284 aAlaGlnAsnGlnProSerGlnLeuProProGlnSerGlnThrGlnPro 300
  |||
948 TTGCTCAGACACCATCACAACCTCCGCGCAGTGGCAGACCCAGCAGCT 997
301 LeuValSerGlnAlaGlnAlaLeuProGlyGlnMetLeuTyrThrGlnPr 317
  |||
998 TTGGTGTCAACAGCGCCAGCAGCTCTCCCTGACAAATGTGTATACCAAC 1047
317 oProLeu 319
  |||
1048 ACCAAT 1054

```

seq\_name: /SID52/gcgdata/geneseq/geneseqn/NA2001.DAT:AAF65986

seq\_documentation\_block:

ID AAF65986 standard; cDNA: 386 BP.

AA65986;

09-APR-2001 (first entry)

Novel human polynucleotide, SEQ ID NO: 1742.

Human: cytostatic; gene therapy; colon cancer; prostate cancer; breast cancer; lung cancer; cancer detection; ss.

Homo sapiens.

MO200102568-A2.

11-JAN-2001.

30-JUN-2000; 2000MO-US18374.

02-JUL-1999; 99US-0142310.

02-JUL-1999; 99US-0142311.

(CHIR) CHIRON CORP.

(HRSE-) HRSEQ INC.

Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;

Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;

Ckenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;

Kita D, Garcia V, Jones LW, Strache-Grain B;

WPI; 2001-091805/10.

Library of polynucleotides for diagnosing a cancerous state of a

mammalian cell and detecting cancer, particularly of the colon or

prostate, comprises 3351 human polynucleotide sequences -

Claim 9; Page 793; 1046pp; English.

The present sequence is one of 3351 sequences in a library of human polynucleotides. The library is used to detect differentially expressed genes correlated with a cancerous state of a mammalian cell and can detect colon, prostate, breast and lung cancer. The library can be used to produce probes for detection of mRNA and to produce additional copies of the polynucleotides. The probes can be used for chromosome mapping of the polynucleotide and for detection of transcription levels. Ribozymes or antisense oligonucleotides can be generated. The polynucleotides and their gene products are used as genetic or biochemical markers (e.g. in blood or tissues) that will detect the earliest changes along the carcinogenesis pathway and/or monitor the efficacy of therapies and preventive interventions. The polynucleotides, polypeptides and antibodies against them can be used in pharmaceutical compositions to treat the cancers and proliferative disorders such as neoplasia, dysplasia and hyperplasia.

Sequence 386 BP; 50 A; 88 C; 138 G; 110 T; 0 other;







XX Probe: microarray; human; placenta; antenatal diagnosis;  
 KM genetic disorder; ss.  
 XX Homo sapiens.  
 OS  
 XX  
 PN W0200157272-A2.  
 PD 09-AUG-2001.  
 PE 30-JAN-2001; 2001WO-US00663.  
 PF 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 DR WPI: 2001-488897/53.  
 XX Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human placenta -  
 DR  
 XX WPI: 2001-488897/53.  
 PT  
 XX  
 PS Claim 25; SEQ ID NO 719; 654pp; English.  
 CC  
 XX The present invention relates to single exon nucleic acid probes (SENPs).  
 CC The present sequence is one such probe. The probes are useful for  
 CC producing a microarray for predicting, measuring and displaying gene  
 CC expression in samples derived from human placenta. The probes are useful  
 CC for antenatal diagnosis of human genetic disorders.  
 CC  
 XX  
 SQ Sequence 436 BP; 105 A; 127 C; 117 G; 87 T; 0 other;

alignment\_scores:  
 Quality: 413.50 Length: 111  
 Ratio: 4.307 Gaps: 1  
 Percent Similarity: 86.486 Percent Identity: 78.378

alignment\_block:  
 US-09-668-119-3 x AA132033 ..  
 Align seg 1/1 to: AA132033 from: 1 to: 436

```

. 95 GlnPrometSerleuSerGlyGlnProProGlyThrSerGlyMetAl 111
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
123 CAGCCGTCGAGCGTTCTGGAACTGCTGCTGAGTGAAGTGC.... 167
111 apRho1SerMetAlaValAlaSerThrAlaThrProGlyThrGlnLeuG 128
168 ..... GACTCTGTCCTTTCTCAGCCAGCTGC 195
128 InLeuGlnAlaValAlaLeuGlnGlnGlnGlnGlnGlnGln 144
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
196 AGCTTCAGCAGGTGGCGCTGCAGCAGCAGCAGCAGCAGTTCAG 245
145 GlnGlnGlnGlnAlaAlaLeuGlnGlnGlnGlnGlnGlnGln 161
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
246 CAGCAGCAGCAGCGCGCTACAGCAGCAGCAGCAGCAGCAGCA 295
296 GCAGTTCAGGCTCAGCAGAGTGCATGACAGCAGCAGTTCAGTAG 345
178 aGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnHisLeu 194
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
346 TGCAGCAGCAGCAGCAGCTCCAGCAGCAGCAGCAGCAGCAGCACTTA 395
  
```

195 IleuLeuHisGlnAsnGlnGlnGlnIle 205  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 396 ATTAATTCATCATCAATAATCAGCAGCAGCTA 428

seq\_name: /SID52/gcgdata/geneseq/geneseqn/NA2001.DAT.AA100701  
 seq\_documentation\_block:  
 ID AA100701 standard; DNA: 436 BP.  
 AC AA100701;  
 DT 09-OCT-2001 (first entry)  
 DE Probe #692 used to measure gene expression in human breast sample.  
 XX  
 DE Probe #692 used to measure gene expression in human breast sample.  
 XX  
 KW Probe: human; breast disease; breast cancer; development disorder; ss;  
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200157270-A2.  
 PD 09-AUG-2001.  
 PE 29-JAN-2001; 2001WO-US00661.  
 PF 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 DR WPI: 2001-476286/51.  
 XX Novel single exon nucleic acid probe used to measuring gene expression  
 PT in a human breast -  
 DR  
 XX  
 PS Claim 25; SEQ ID NO 692; 322pp; English.  
 CC  
 XX The present invention relates to novel single exon nucleic acid probes.  
 CC The present sequence is one such probe. The probes are useful for  
 CC measuring human gene expression in a human breast sample, where the probe  
 CC hybridises at high stringency to a nucleic acid expressed in the human  
 CC breast. The probes are useful for predicting, diagnosing, grading,  
 CC staging, monitoring and prognosing diseases of the human breast,  
 CC particularly those diseases with polygenic aetiology. The diseases  
 CC include: breast cancer, disorders of development, inflammatory diseases  
 CC of the breast, fibrocystic changes, proliferative breast disease and  
 CC non-carcinoma tumours.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WPI  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SQ Sequence 436 BP; 105 A; 127 C; 117 G; 87 T; 0 other;

alignment\_scores:  
 Quality: 413.50 Length: 111  
 Ratio: 4.307 Gaps: 1  
 Percent Similarity: 86.486 Percent Identity: 78.378

alignment\_block:  
 US-09-668-119-3 x AA100701 ..  
 Align seg 1/1 to: AA100701 from: 1 to: 436



```

95 GlnPrometSerLeuSerGlyGlnProProGlyThrSerGlyMetAl 111
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
123 CAGCCCTCAGCAGCTTCTGGGAATCTGCGCTGAGTCACTGCGC.... 167
111 aProHisSerMetAlValValSerThrAlaThrProGlnThrGlnLeu 128
168 .....GACTCGGCTCTTTCTCCAGCCCAAGCTGC 195
128 LnuGlnGlnValAlaLeuGlnGlnGlnGlnGlnGlnGlnPheGln 144
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
196 AGCTCCAGCAGAGTGGCTGCTGAGCAGCAGCAGCAACAGCAGCTTCAG 245
145 GlnGlnGlnGlnAlaAlaLeuGlnGlnGlnGlnGlnGlnGlnGln 161
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
246 CAGCAGCAGCAGCAGCGCGCTGACAGCAGCAGCAGCAGCAACAGCA 295
161 nGlnPheGlnAlaGlnGlnSerAlaMetGlnGlnGlnPheGlnAlaVal 178
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
296 GCAGTTCAGGCTCAGCAGAGTGCATGACAGCAGCTTCCAAACAGTAG 345
178 alGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 194
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
346 TGCAGCAGCAGCAGCAGCGCTCCAGCAGCAGCAGCAGCAGCAGCAGTA 395
195 IleLysLeuHisGlnHisGlnHisGlnGlnGlnGlnGlnGln 205
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
396 ATTAATTGCTCATCTCAAAATCAGCAGAGGTA 428

```

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAI20017

seq\_documentation\_block:  
ID AAI20017 standard; LNA; 316 BP.

AAI20017;

12-OCT-2001 (first entry)

Probe #9950 for gene expression analysis in human cervical cell sample.

Probe; human; microarray; gene expression; cervical epithelial cell;  
cervical cancer; ss.

Homo sapiens.

WO200157278-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00670.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632366.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-488901/53.

Human genome-derived single exon nucleic acid probes useful for  
analyzing gene expression in human cervical epithelial cells -

Claim 25; SEQ ID NO 9950; 487bp; English.

The present invention relates to human single exon nucleic acid probes  
(SENP). The present sequence is one such probe. The SENPs are derived  
from human HeLa cells. The SENPs can be used to produce a single exon  
microarray, which can be used for measuring human gene expression in a

CC sample derived from human cervical epithelial cells. By measuring gene  
CC expression, the probes are therefore useful in grading and/or staging  
CC of diseases of the cervix, notably cervical cancer.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 316 BP; 81 A; 104 C; 87 G; 44 T; 0 other;

alignment\_scores:

Quality:	393.00	Length:	80
Ratio:	4.912	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	98.750

alignment\_block:  
US-09-668-119-3 x AAI20017

Align seg 1/1 to: AAI20017 from: 1 to: 316

```

126 GlnPheGlnLeuGlnGlnValAlaLeuGlnGlnGlnGlnGlnGln 142
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3 CAGCTCAGCTCCAGCAGAGTGGCTGCTGAGCAGCAGCAGCAACAGCA 52
142 nPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 159
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
53 GTTCCAGCAGCAGCAGCAGCGCGCTGACAGCAGCAGCAGCAGCAGC 102
159 LngGlnGlnPheGlnAlaGlnGlnSerAlaMetGlnGlnGlnPheGln 175
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
103 AACAGCAGCAGTTCAGGCTCAGCAGAGTGCATGACAGCAGCAGTTC 152
176 AlaValAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 192
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
153 GCAGTAGTCAGCAGCAGCAGCAGCTCCAGCAGCAGCAGCAGCAGCA 202
192 nHisLeuIleLysLeuHisGlnHisGlnHisGlnGlnGlnGln 205
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
203 GCATCTAATTAAATTCATCATCAAAATCAGCAGAGGTA 242

```

seq\_name: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:AAI45215

seq\_documentation\_block:  
ID AAI45215 standard; DNA; 316 BP.

AAI45215;

17-OCT-2001 (first entry)

Probe #13901 used to measure gene expression in human placenta sample.

Probe; microarray; human; placenta; antenatal diagnosis;  
genetic disorder; ss.

Homo sapiens.

WO200157272-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US00663.

04-FEB-2000; 2000US-0180312.

26-MAY-2000; 2000US-0207456.

30-JUN-2000; 2000US-0608408.

03-AUG-2000; 2000US-0632366.

21-SEP-2000; 2000US-0234687.

27-SEP-2000; 2000US-0236359.

04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;



[illegible]



OS	Homo sapiens.
XX	
KW	Polycombtarget 1; hphl gene; oncogene; human; neoplasia;
CC	dysplasia; hyperplasia; diagnosis; tumour; metastasis; therapy; ss.
XX	
FH	
FT	Key
FT	CDS
XX	
XX	Location/Qualifiers
XX	1..3015
XX	/+tag- a
XX	
XX	W09807860-A1.
XX	
XX	26-FEB-1998.
XX	
XX	
XX	
XX	22-AUG-1997; 97WO-US14886.
XX	
PR	06-FEB-1997; 97US-0036939.
PR	23-AUG-1996; 96US-002349.
XX	
XX	04-DEC-1996; 96US-0031569.
XX	
PA	(CHIR ) CHIRON CORP.
XX	
FL	Randazzo F;
DR	WP1; 1998-169164/15.
DR	P-P5DB; AAW37856.
XX	
PT	Isolated human polycombtarget 1 oncogene - used to develop products
PT	for diagnosis and therapy of proliferative and developmental
PT	disorders, e.g. neoplasia, dysplasia or hyperplasia
XX	
XX	Claim 7; Page 25-27; 39pp; English.
XX	
XX	This polynucleotide comprises a novel human oncogene termed
CC	polycombtarget 1 (hphl) that is implicated in neoplastic disorders,
CC	it codes for a 1004-amino acid polypeptide (see AAW37856). The gene
CC	maps to human chromosome 12p13. Also claimed are: (1) an hphl 3'
CC	polypeptide consisting of at least 22 contiguous amino acids; (2)
CC	an hphl fusion protein; (3) antibodies which specifically bind to a
CC	portion of hphl protein; (3') a homologously recombinant cell having
CC	a new transcription initiation unit (TIV), which comprises in 5' to
CC	3' order: (a) an exogenously regulatory sequence; (b) an exogenous
CC	exon, and (c) a splice donor site, where the TIV is located
CC	upstream to a coding sequence of an hphl gene, where the exogenous
CC	regulatory sequence controls transcription of the coding sequence of
CC	the hphl gene; (6) a method of identifying neoplastic tissue of a
CC	human, where over-expression of the hphl gene identifies the tissue
CC	as being neoplastic; (7) a method to aid in the diagnosis or
CC	prognosis of neoplasia, where a difference in the hphl gene, mRNA,
CC	or protein between a first and second tissues indicates neoplasia
CC	in the first tissue; (8) a method to aid in detecting a genetic
CC	predisposition to neoplasia; (9) a method of identifying a human
CC	chromosome 12; and (10) a method of inducing a cell to
CC	differentiate by contact with a hphl gene or expression product.
CC	The products can also be used to inhibit hphl expression to
CC	suppress neoplasia, dysplasia, or hyperplastic cell growth. They
CC	can be used to treat e.g. tumours, ankytic hereditary ectodermal
CC	dysplasia, congenital alveolar dysplasia, epithelial dysplasia of
CC	the cervix, fibrous dysplasia of bone, or mammary dysplasia,
CC	endometrial, adrenal, breast, prostate, or thyroid hyperplasia or
CC	pseudo-epitheliomatous hyperplasia of the skin. Even in disorders
CC	in which hphl mutations are not implicated, down-regulation or
CC	inhibition of hphl expression can have therapeutic application.
CC	In these disorders, decreasing hphl expression can help to suppress
CC	tumours. Similarly, in tumours where hphl expression is not
CC	aberrant, effecting hphl down-regulation can suppress metastases.
CC	Antisense polynucleotides can also be used to induce differentiation
CC	of a progenitor cell, e.g. in order to study the process of
CC	differentiation and test compounds which affect this process.
XX	
XQ	Sequence 3879 BP; 967 A; 1108 C; 954 G; 850 T; 0 other;



```

315  rClnProProLeuLysPheValArgAlaProMetVal..... 327
1444  CAAACACCTGTAGCACCATTATCAAGCCGCTCAGTTAGGGGCGCCGTAAGA 1543
338  .....ValGlnGlnProProValGlnProGlnValGlnGlnGln 341
1544  TGTCAGTGCCTCCAGCAACCAACA.....CCCATATCCTGTGTCAAGTT 1587
342  ThrAlaValGlnThrAlaGlnAlaAlaGlnMetValAlaProGlyValG1 358
1588  GTAGGCACACTCGACAGCAGCAGTACAGCCACAGCAGCAGAGCTTTGGGTTGGC 1637
358  nValSerGlnSerSerLeuProMetLeuSerSerSerProSerProGlyGln 375
1638  ACAGCTGGCAGGCTGCTGTACT..... 1659
375  nValGlnThrProGlnSerMetProProProGlnProSerProGln 391
1660  .....ACTTCCCGCGGGATGCCAGGTACAGTGCAG...TCTGGTCAG 1698
392  ProGlyGlnProSerSerGlnProAnSerAsnValSerSerGlyProAl 408
1699  GCCCATTTGGCCTCTCGCCACCTTATCCAGGCTCTGGTGACTGCA 1748
408  pProSerProSerSerPheLeuProSerProSerProGlnProSerGln 424
1749  GGAGTGCCTCCCAACATTGGCCCTGGGATGACCCCTGCTCCTGTGCAGG 1798
425  .....SerProValThr 428
1799  GCACACGACATGTGTAAAGGGTGGGGCTACCACTCCTCATTCTGTGTGA 1848
429  AlaArgThrProGlnAnPheSerValProSerProGlyProLeuAnThr 445
1849  GCCCAGGTCCCTGCTGCTCTTCTATATGCACT.....GTGCACATT 1888
445  rProValAnProSerSerValMetSerProAlaGlySerSerGlnAlaG 462
1890  GCCGGGTAAACCCCAACACATTG.....GCTGTCAACAGCAAGGCTG 1930
462  lngLnglnGlnThrLeuAspPheLeuGlnLeuSerLysTyrIleGlu 478
1931  ACTCTGAGAGGAGAGATGATGTCACACATTGGGTTCATATCCTTCCCT 1980
479  ProLeuArgArgMetIleAsnLysIleAspLysAsnGlnAspArgLysLy 495
1981  GCCAAGCATCTTCCAGTAGCAGAAAGCCCAAAAGTCATGACGAGAGAGAG 2033
495  sAspLeu...SerLysMetLysSerLeuLeuAspIle..... 506
2031  CAGTCTTGAGAAAGAAAGCTCAATCAGTGGGTAAATGATGATCTAATACTC 2080
507  .....LeuThrAspProSerLysArgCysProLeu 516
2081  CAAGCAGTAGAATAGTACCTTGACC...CCCGCCCTTCAGTACCGCCT 2127
517  LysThrLeuGlnLysCysGluIleAlaLeuGlnLysLeuLysAsnAspM 533
2128  CCTACACTGAGCCATGGTGT.....CTAGACAAATGGGTGACTCAAA 2166
533  eIArgCysProLeuProIHisArgPro 541
2169  AC.....CCCAACAGGCA 2182
seq_name: /SIDS2/gcgcdata/geneseq/geneseqn/NA1998.DAT.AAV21060
seq_documentation_block:
ID AAV21060 standard; cDNA; 3879 BP.
AC AAV21060;
XX
XX 09-JUL-1998 (first entry)

```

```

XX  cdna encoding the a tumour suppressor gene called polyhomeotic 1.
DE
XX
XX  Tumour suppressor gene; polyhomeotic 1; hph1: human chromosome 12p13;
KM  identification; neoplastic tissue; cellular differentiation; diagnosis;
KW  neoplasia; ss.
XX
XX  Homo sapiens.
OS
PH  Key
FT  CDS
XX  Location/Qualifiers
XX  1..3015
XX  /*tag= a
XX
XX  WO9807858-A1.
XX
XX  26-FEB-1998.
XX
XX  22-AUG-1997; 97WO-US14866.
XX
XX  06-FEB-1997; 97US-0036939.
XX  23-AUG-1996; 96US-0024349.
XX  04-DEC-1996; 96US-0031569.
XX
XX  (CHIR ) CHIRON CORP.
XX
XX  Randazzo F;
XX  WPI: 1998-169162/15.
XX  DR  P-PSDB; AAW52830.
XX
XX  Isolated human polyhomeotic 1 oncogene - used to develop products
PT  for diagnosis and therapy of proliferative and developmental
PT  disorders, e.g. neoplasia, dysplasia or hyperplasia
XX
XX  Claim 5; Pages 25-27; 40pp; English.
XX
XX  The present sequence encodes a novel human tumour suppressor gene
CC  termed polyhomeotic 1 (hph1). The hph1 gene maps to human chromosome
CC  12p13, a region which is frequently lost in non-small cell lung cancer
CC  and breast cancer. A method of identifying neoplastic tissue of a
CC  human comprises comparing the expression of a hph1 gene in a tissue of
CC  a human suspected of being neoplastic with the expression of a hph1
CC  in a tissue of the human which is normal. Under-expression of the hph1
CC  gene identifies the subject as having neoplastic tissue. The hph1
CC  oncogene functions to suppress neoplasia and dysplastic or hyperplastic
CC  cell growth as well as to induce cellular differentiation. The cdna,
CC  protein and vectors can be used as diagnostic and therapeutic tools for
CC  proliferative and developmental disorders and to identify a p13 region
CC  of a human chromosome 12. They can be used for the detection, diagnosis
CC  or prognosis of neoplasia or for detecting a genetic predisposition to
CC  neoplasia. They can also be used to treat tumours.
XX
XX  Sequence 3879 BP; 965 A; 1108 C; 954 G; 850 T; 2 other;
XX
XX
XX  alignment_scores:
XX  Quality: 384.00 Length: 526
XX  Ratio: 1.367 Gaps: 23
XX  Percent Similarity: 53.422 Percent Identity: 28.707
XX
XX  alignment_block:
XX  US-09-668-119-3 x AAV21060 ..
XX
XX  Align seg 1/1 to: AAV21060 from: 1 to: 3879
XX
XX  60 GlnSerLeuThrGlyGlyProAlaAlaGlyAlaAlaGlyTleGlyMePr 76
XX  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX  778 CAGTCTCCACCTAGTCAAGTGTGTGAAGGAGGAGGAAATGACATCCC 827
XX  ProArgGlyProGlyGlnSerLeuGlyGlyMetGlyS 89
XX  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX  828 AGGGTCCATGGGTCCAGGTGGAGGTGGCAGGACATGTTGGTTGGGTC 877

```



```

89  erPheGlyAlaMetGlyGlnProMetSerLeuSerGlyGlnProPro 105
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
878  AGTTCCTCTCCAGTA.....ATGGGTGGGAGGCTGTCACGAGAG 921
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
106  GlyThrSerGlyMetAlaProHisSerMetAlaValSerThrAlaTh 122
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
922  GGTACAGAGGTGTGAGCCCTTGCTGCGAGCCCAACAGTACGTAG 971
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
122  rProGlnThrGlnLeuGlnLeuGlnValAlaLeuGlnGlnGlnG 139
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
972  CCAGGCGACGACGACGAGGCGAAGGTGACGACCAAGAGGACGAG 1021
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
139  GlnGlnGlnPheGlnGlnGlnGlnAlaAlaLeuGlnGlnGln 155
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
1022  CAGATGGAGGTGGCCGACGAGATGTGGGCAATGACCTGACAGGAC 1071
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
156  GlnGlnGlnGlnGlnGlnGlnPheGlnAlaGlnGlnSerAlaMetG 172
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
1072  ACACCTGGGCGCCAGCCAGCACTTATAGCTCAGGCACTTACACAGAT 1121
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
172  nGlnPheGlnAlaValGlnGlnGlnGlnGlnGlnGlnGlnGln 189
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1122  CCAGGCCCATTCACGTATTCAGCAACACAGCAATCCCTCCAGCAGA 1171
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
189  GlnGlnGln.....GlnHisLeuLeuLeuHisHisGlnGlnGln 203
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
1172  AACAGGTGTATCTCAGCAGCATTTGCCATCCACAC.....CAGCAG 1215
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
204  GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnG 219
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1216  CAGTTCACGACCGGAGTCCAGTCCCTTCACACAGCTACACACCTCA 1285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
219  nLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 236
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1266  GTTGGCGACGACGACGACGACCAACACAGCAGCAGCAGCAGCAG 1315
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
236  In.....AlaLeuGlnAlaGlnProPheGlnGlnGlnProPro 248
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
1316  AGCCCGAAGCCACACACCTCCTAGCTGCTCCAGCCACAGCTCCACCT 1365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
249  MetGlnProGlnProPheProPheSerGlnAlaLeuProGlnGln 265
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1366  ACTCAG.....CAGCTCCACCTCCACCTCCAGTCC.....CA 1394
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
265  nGlnMetHisHisThrGlnHisHisGlnProProProGlnProGln 282
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1395  GCAGCAAGCCCAACCTGCTGCTCAG..... 1422
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
282  rProProValAlaGlnAlaGlnProSerGlnLeuProProGlnSer 298
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1423  ..CCCATGCTTCAGTCTTACCTTGTCTTCCACCTGATGACGCCCT 1470
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
299  GlnProLeuValSerGlnAlaGlnAlaLeuProGlnGlnMetLeu 315
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
1471  AAGCCCAACCAATCC.....ATCCAAATC 1493
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
315  rGlnProProLeuLysPheValArgAlaProMetVal..... 327
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
1494  CAACCCACCTGTAGCTCCTATCAAGCCGCTCAGTTAGGGCCCTAGA 1543
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
328  .....ValGlnGlnProProValGlnProGlnValGlnGlnGln 341
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
1544  TGTACAGTGGCCAGCAACCA.....CCCATATCCCTGTGCAAGTT 1587
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
342  ThrAlaValGlnThrAlaGlnAlaAlaGlnMetValAlaProGlyVal 358
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
1588  GTAGGCACTGCACAGCAGGTACAGCCAGGCAAGGCTTGGGGTGGC 1637
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
358  nValSerGlnSerSerLeuProMetLeuSerSerProSerProGln 375
    ::::: ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
1638  ACAGTGGCAGCTGTGTACT..... 1659
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
375  InValGlnThrProGlnSerMetProProProGlnProSerProGln 391
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

1660  ....ACTTCCCGGGGATGCCAGTACAGTACAG...TCTGGTCA 1698
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
392  ProGlyGlnProSerSerGlnProAsnSerAsnValSerGlyProAl 408
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1699  GCCCATTTGGCCCTTCGCCACCTTCATCCACAGCTCCTGCTGCTCA 1748
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
408  aProSerProSerSerPheLeuProSerProSerProGlnProSer 424
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1749  GGAATGCCCTCCACATTTGGCCCTGGATGACCTTGTCTGTGCTGAG 1798
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
425  .....SerProValThr 428
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1799  GGACAGCAGATGTGTAAAGGTGGGCTACACCTCCTCAGCTGTGTA 1848
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
429  AlaArgThrProGlnAsnPheSerValProSerProGlyProLeuAs 445
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1849  GCCCAGTCCCTGCTGCTCTTATGACAGTCT.....GTGCACTT 1889
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
445  rProValAsnProSerSerValMetSerProAlaGlySerSerGlnAla 462
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1890  GCCGGGTAAACCCAGACATTTG.....GCTGTCAACGCAAGGCTG 1930
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
462  LucGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 478
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1931  ACTGTGAGGAGAGAGATGATGTCTCAGATTTGGGTTCAATGCTTCT 1980
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
479  ProLeuArgArgMetLeuAsnLysLysLysLysLysLysLysLys 495
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1981  GCCAAGCATCTCCAGTACAGAAAGCCCAAGTATGACGAGAGAGAG 2030
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
495  SasPheLysSerLysMetLysSerLeuLeuAspLys..... 506
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2031  CAGTCTTGAGAAAGCTGATCATGAGTGAATGATGCTAATATCTC 2080
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
507  .....LeuThrAspProSerLysArgCysProLeu 516
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2081  CAAGCAGTGAATAGTACCTTGACC...CCGCCCCCTTCAGTACGGCT 2127
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
517  LysThrLeuGlnLysCysGlnLeuLeuLeuGlnLysLysLysLys 533
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2128  CCACTACCTACCTACCTGCTG.....CTAGCAAAATGGGTGACTCAA 2168
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
533  eArgCysProLeuProHisArgPro 541
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2169  AC.....CCCAACAGGCCA 2182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV22682
seq_documentation_block:
ID  AAV22682 standard; DNA; 2214 BP.
XX  AAV22682;
XX  28-JUL-1998 (first entry)
XX  DE  New DNA sequence isolated from Pinctada fucata.
XX  Pinctada fucata; protein; cosmetic; ds.
XX  Pinctada fucata.
XX  JP10080285-A.
XX  31-MAR-1998.
XX  PD  28-MAY-1997; 97JP-0138461.
XX  PF  15-JUL-1996; 96JP-0184459.
XX  PR  (MTK-) MIKIMOTO SETYAKU KK.
XX  PA  WPI; 1998-254410/23.
XX  DR

```



XX New cDNA and e.g. vector, host cell and polypeptide - used to  
 PT produce polypeptide in high yields, which is used in cosmetics  
 XX  
 PS Claim 1; Page 7; 15pp: Japanese.  
 XX

CC The present sequence represents a new DNA sequence isolated from *Pinus*  
 CC *fucata*. The polypeptide be used as an ingredient in cosmetics.  
 XX

SQ Sequence 2214 BP; 393 A; 451 C; 891 G; 479 T; 0 other;

# alignment\_scores:

Quality: 381.50 Length: 597  
 Ratio: 1.408 Gaps: 25  
 Percent Similarity: 45.394 Percent Identity: 25.628

## alignment\_block:

US-09-668-119-3 x AAV22682/rev ..

Align seg 1/1 to reverse of: AAV22682 from: 1 to: 2214

```

50 AlaSerValSerAPro.....MetAsnAlaLeuG1 60
   |||.....|||
1631 GCGGCAGCAGGAGATCCACCTCCACGCACTTTGCTTCTCAAGCCCTCT 1582
   |||
60 nSerLeuThyGlyPro.....AlaAlaGlyAla 71
   |||
1581 AAGTGTCTTCTCCACCTGATGACAGCAGCGGCGCTGCACAG 1532
   |||
71 IagIyIleGlyMetProIroArgIyProGlyInserLeuGlyMet 87
   |||
1531 CTGGGCGAGATCCCGACCAAGACCGCT..... 1503
   |||
88 GlySerPheGlyAlaMetGlyGlnProMetSerLeuSerGlyGlnPro 104
   |||
1502 .....AGTCTCCAAAGTCCACCAAAACCTCCTCC 1474
   |||
104 oProGlyThySerGlyMetAlaProHisSerMetAlaValSerThra 121
   |||
1473 ACCT.....GCACCAGCAGCAGCTAGTGCAGCGGCTAAAG 1439
   |||
121 IatHPro.GlnThrGlnLeuGlnLeuGlnAlaAlaLeuGlnGln 137
   |||
1438 CTCCGCCCTCCCTCCACCTCCAAAGTCCACCTCCGCTCCGACAGCG 1389
   |||
137 nGlnGlnGlnGlnInPheGlnGlnGln..... 147
   |||
1388 AGAGCAGCAGCTAGAGCTCCAGCTCCACCGCGCTCCTCCTCCGTTACC 1339
   |||
148 .....GlnAlaAlaLeuGlnGln 154
   |||
1338 ACCTCCGTTACACACCGCTCCACCTCCAGCTGGAAGCTGATGACAGATG 1289
   |||
155 GlnGlnGlnGlnGlnGlnGlnInPheGlnAlaGln..... 166
   |||
1288 CAGATGCTGAAGCAGATGCTGAGCTTTCAATAGACCTGCAGATCCTTT 1239
   |||
167 GlnSerAlaMetGlnGlnGlnInPheGlnAlaVal..... 178
   |||
1238 AAGACATCCTCAATATGTCATTTGAGATTTTGTGCTGTCTACTGCGGA 1189
   |||
178 ..... 178
   |||
1188 TGCTTTACAGAGGCTTTTGACAGAGCTCATCTCCTGAGAGGTCCTT 1139
   |||
178 ..... 178
   |||
1138 TAATGTCTCTAATGCTCCCTTATACACATCTTGTCTGAAGCTGAA 1089
   |||
179 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln..... 192
   |||
1088 GCAGATGTCAGCAGCGGCTGACAGCGGACGACGACGACGCGCTCCGCC 1039

```

```

193 .....HisLeuIleLysLeuHisGlnAsnGln 203
   |||
1038 TCTCTCTCCACCACTCCACCTCTCCGCCACACGAGGCTGGGGG 989
   |||
204 GlnIleGlnGlnGlnGln.....LeuG1 213
   |||
988 CAGCTCAGCAGCGGACGACGACGCTGCTGCTCTCTCTCTGCA 939
   |||
213 nArgIleAlaGlnLeuGlnGlnGlnGln..... 224
   |||
938 CCGCGTCTACCTCTCTCCACCTGTAAGAACCAACACGACGATCTCC 889
   |||
225 .....GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnAlaLeu 238
   |||
888 ACCTCGGACGACGCGGCGGACGACGACGACGACGACGACGATCCTC 839
   |||
838 CATAT...CTCCGAGGCCACCAAGACCTCCGAGGCTCCAAATCCGCC 793
   |||
255 oProSerGlnAlaLeuProGlnGlnGlnGlnGlnGlnHisIstHrGln 272
   |||
792 ACCGAGTCTCTCA..... 780
   |||
272 IHisGlnProProGlnProGlnGlnProProValAlaGlnAsnGln 288
   |||
779 .....AGTCCGCCACCGAGTCTCCAAATCCACCTGCG..... 747
   |||
289 ProSerGlnLeuProProGlnGlnGlnGlnGlnGlnGlnGlnGln 305
   |||
746 .....CTCTCCAGCGGCGCA.....GCGGCTGGGC 718
   |||
305 aGlnAlaLeuProGlyGlnMetLeuThyThrGlnProProLeuLysPhe 322
   |||
717 TGCAGCAGCAGCTCTCCAAATCTACCTCGGCTCTCTGACGCGCAG 668
   |||
322 aAlaArgAlaProMetValGlnGlnGlnGlnGlnGlnGlnGlnGln 338
   |||
667 CACCGGCTGGGCTGGGCGACGCTCCACCAATCTCACCCCGGCG... 621
   |||
339 GlnGlnGlnThraValGlnThraGlnAlaGlnAlaGlnMetValAla 355
   |||
620 .....GACGACGCGGAGCTGTCAGCAGCAGCAGCAGCAGCTCC 577
   |||
355 oGlyValGlnValSerGlnSerLeuProMetLeuSerProSerP 372
   |||
576 ACCAGCTCCACCGAGGCGAAGAGCTGCACCAAGA...TCATGAATCTA 530
   |||
372 roGlyGlnGlnValGlnThrProGlnSerMetPro...ProProGln 387
   |||
529 AATCAATTAATATGCTCTCCACCGCAAGTCCGCGCAGACCTCAAGTCCG 480
   |||
388 ProSerProGlnProGlnProSerSerGlnProAsnSerAsnValSe 404
   |||
479 CCAAGTCCACCTCCGAGACCTCG..... 456
   |||
404 rSerGlyProAlaProSerProSerSerPheLeuProSerProSerPro 421
   |||
455 .....CCCAATCCAAATCCGAGT.....CCAGCAGCTGCTCCG 422
   |||
421 InProSerGlnSerProValThraIArgThrProGlnAsnPheSerVal 437
   |||
421 CTCCTGGCTGCTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 372
   |||
438 ProSerProGlyProLeuAsnThrProValAsnProSerSerValMetSe 454
   |||
371 CAGACACACCTCCGCGGCGCTCCATTCACCTCATTCATCCAGTCC 322
   |||
454 rProAlaGlySerSer.GlnAlaGlnGlnGlnGlnGlnGlnGlnGln 470
   |||
321 CCATTCATCATGTCATGCTGTCGTCATCATCATCATTCATTCACATTC 272

```



```

471 LysGlnLeuSerLysTyrIleGluProLeuArgMetIleAsnLysII 487
271 CACCCC.....CGCGCGCGCGTCCGCG..... 251
487 GAspLysAsnGlnAspArgLysLysAspLeuSerLysMetLysSerLeuL 504
250 .....CGCGCGCTCC 241
504 euAspIleLeuThrAspProSerLysArgCysProLeuLysThrLeuGln 520
240 PCGAGCAGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 197
521 LysCysGluIleAlaLeuGluLysLeuLysAsnAspMetArgCysProLe 537
196 .....CGTAAAGAGCCTGCACCTGCTCAT 174
537 uPro.....HisArgProArgCysHisArgPro 546
173 GCCGAAACGCAACCAATGCCCCATGATCATTAACCA 137
seq_name: /SIDIS2/gcgdata/geneseq/geneseqn/NA1998.DAT:AAV22683
seq_documentation_block:
ID AAV22683 standard; cDNA to mRNA; 3331 BP.
XX AAV22683;
XX
XX 28-JUL-1998 (first entry)
XX
XX DE New DNA sequence isolated from Pinctada fucata.
XX
XX KW Pinctada fucata; protein; cosmetic; ds.
XX
XX OS Pinctada fucata.
XX
XX FH Key! Location/Qualifiers
XX FT CDS 50..2266
XX
XX JPI0080285-A.
XX
XX PD 31-MAR-1998.
XX
XX PF 28-MAY-1997; 97JP-0138461.
XX
XX PR 15-JUL-1996; 96JP-0184459.
XX
XX PA (MIKI-) MIKIMOTO SEIYAKU KK.
XX
XX DR WPI: 1998-254410/23.
XX
XX DR E-PSDB; AAW56163.
XX
XX PT New cDNA and e.g. vector, host cell and polypeptide - used to
XX PT produce polypeptide in high yields, which is used in cosmetics
XX PS Claim 2; Pages 7-9; 15pp; Japanese.
XX
XX CC The present sequence represents a new DNA sequence isolated from Pinctada
XX CC fucata. The encoded polypeptide be used as an ingredient in cosmetics.
XX
XX SQ Sequence 3331 BP; 762 A; 701 C; 1022 G; 846 T; 0 other;
XX
alignment scores:
Quality: 381.50 Length: 597
Ratio: 1.408 Gaps: 25
Percent Similarity: 45.394 Percent Identity: 25.628
alignment block:
us-09-668-119-3 x AAV22683/rev
Align seq 1/1 to reverse of: AAV22683 from: 1 to: 3331

```

```

50 AlaSerValSerAspPro.....MetAsnAlaLeuGI 60
1680 GCGGACAGCGAGATCCACCTCCAGCAGCATTTGCTTCTCAAAAGCCTCT 1631
60 uSerLeuThrGlyGlyPro.....AlaIleGlyAla 71
1630 AAGTCCTCTCTCCACACCTGATGACAGCGGAGCGGCGTGCAGCAG 1581
71 IagIyIleGlyMetProProArgGlyProGlyGlnSerLeuGlyGlyMet 87
1580 CTGCGGAGATGCCCGACCAAGACCGCT..... 1552
88 GlySerPheGlyAlaMetGlyGlnPrometSerLeuSerGlyGlnProPr 104
1551 .....AGTCCTCCAGTCCACCAAAACCTCTCC 1523
104 oProGlyThrSerGlyMetAlaProHisSerMetAlaValAlaSerThra 121
1522 ACCT.....GCACCAAGCAGCAGCTAGTGCAGCGGCGTGAAG 1488
121 IeThrPro.GlnThrGlnLeuGlnLeuGlnGlnAlaAlaLeuGlnGln 137
1487 CTGCGGCTCGGCTCCACCTCCAGTCCAGTCCGCGCTCTGACAGCG 1438
137 nGlnGlnGlnGlnGlnPheGlnGlnGln..... 147
1437 AGAGCAGCAGCTAGAGCTCAGCTCCACCGCGCGCTCTCTCCCTTAC 1388
148 .....GlnAlaAlaLeuGlnGln 154
1387 ACCTCGTTACACACCGCGCTCCACCTCCAGTCCAGTGAAGTGAAGATG 1338
155 GlnGlnGlnGlnGlnGlnGlnPheGlnAlaGln..... 166
1337 CAGATGCTGAAGCAGATGCTGAGCTTTCAATGACCTGCAGATCCTT 1288
167 .GlnSerAlaMetGlnGlnPheGlnAlaValAla..... 178
1287 AAGACATCCTCAAAATGCTCAATTGAGATTTGTGCTGACTAGTGA 1238
178 ..... 178
1237 TGCTTTAGCAGAGGCTTTTCAGAGGCTCCATTACTCTGAGAAAGTCTT 1188
178 ..... 178
1187 TAATGCTCTTAATGCGCTTATACCACTGATTTGCTAGAGCTAA 1138
179 .GlnGlnGlnGlnGlnLeuGlnGlnGlnGlnGlnGln..... 192
1137 GCAGATGACAGCAGCGGCTCCAGCGGAGCGAGCAGCAGCGCGCTCCGC 1088
193 .....HisLeuIleLysLeuHisGlnAlaGln 203
1087 TCCTCTCCACCACTCCACCTCCCGCCACACCAAGCGGCTGGCGG 1038
204 GlnIleGlnGlnGlnGlnGln.....LeuGI 213
1037 CAGTGCAGCAGCGGAGCAGCAGCAGCTTCGCGCTCTCTCTCGA 988
213 nArgIleAlaGlnLeuGlnLeuGlnGlnGln..... 224
987 CCGGCTCTACCTCTCTGTCACACCTAGAAACCAACACACGAGCTCC 938
225 .....GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnAlaLeu 238
937 ACCTCGCGAGCAGCGGCGGAGCAGCAGCAGCGGAGCAGCAGCTCTC 888
239 GluAlaGlnProProIleGlnGlnProProMetGlnGlnProGlnPro 255
887 CATAT.....CTTCGAGGACCAAGACCTCCGAGGCTCCAAATGCTCCG 842
255 oProSerGlnAlaLeuProGlnGlnLeuGlnGlnMetHisHisThrGln 272

```



```

841 ACCGAGTCTCTCCA..... 829
272 IShISglInProPProglInProGlnInProProValAlaIasInsgIn 288
828 .....AGTCCGGCACCGAGTCTCTCCAAGTCCACCTGGC..... 796
289 ProSerGlnLeuProProglInSerGlnThrGlnProLeuValSerGlnAl 305
795 .....CTCTCTGACGGGAGCA.....GGCGTGGCGG 767
305 agInAlaLeuProGlyGlnMetLeuThrGlnProProLeuLysPheV 322
766 TGAGAGCAGCAGCTCTCCAAGTCTACCTGGCTCTCTCTGACGGGAG 717
322 alArGlaProMetValValGlnInProProValGlnProGlnValGln 338
716 CACGGGTGGCGTGGGGGAGCAGCTCCACCAATCCACCCCGGG... 670
339 GlnGlnGlnThrAlaValGlnThrAlaGlnAlaGlnMetValAlaPr 355
669 .....GCACACCGGAGCTGTGAGAGAGAGAGCAGCAGCTCCAGCTC 626
355 ogIValGlnValSerGlnSerSerLeuProMetLeuSerSerProSerP 372
625 ACCAGCTCCACGAGGAGCAAGAGCTCCACCAAGA...TCATCGAATCTA 579
372 roGlyGlnGlnValGlnThrProGlnSerMetPro...ProPProGln 387
578 AATCAATTAATATGCTGCCACCGCAAGTCCGCCAAGACCTCCAACTCCG 529
388 ProSerProGlnProGlyGlnProSerSerGlnProAsnSerAsnValSe 404
528 CCAGTCCACTCTCGAGACTCCG..... 505
404 rSerGlyProAlaProSerProSerSerPheLeuProSerProSerProG 421
504 .....CCCAATCCAGTCCGAGT.....CCAGACCTGCTCTG 471
421 InProSerGlnSerProValThAlaArgThrProGlnAsnPheSerVal 437
470 CTCCTGGCTGTGCTCCAGCACAGCACACACACCTCCGCGACACACA 421
438 ProSerProGlyProLeuAsnThrProValAsnProSerSerValMetSe 454
420 CCAGCACCTCCTCGGGCGGACCTCCATTCACCGTCATCATCCACAGTC 371
454 rProAlaGlySerSerGlnAlaGlnGlnGlnInThrLeuAsnLysLeu 470
370 CATTCATCATGTCATCGCTCGTGCATCATATGTCATATTCCAATCTC 321
471 LysGlnLeuSerLysThrIleGlnProLeuAlaArgMetIleAsnLysII 487
320 CACCGC.....CGCGCGCGCTCCGC..... 300
487 eaSPlysAsnGlnuSPArgLysLysAsnLeuSerLysMetLysSerLeu 504
299 .....CGTGGGTCC 290
504 euAsnIleLeuThrAspProSerLysArgCysProLeuLysThrLeuGln 520
289 TCCAGCACCGCAGCTCCCGCAGCGGCTCGCGCCCTCCGACAC..... 246
521 LysCysGlnIleAlaLeuGlnLysLeuLysAsnSpmLeuArgCysProLe 537
245 .....CGTAAAGACCTGCACCTGCTCAT 223
537 uPro.....HisArgProArgCysHisArgPro 546
222 GCGGAAGCAGCAACGACATGCCCATGATCATTAACCA 186
seq_name: /SIDS2/gcgdata/geneseg/genesegN/NA1999.DAT.NA132021

```

```
seq_documentation_block:
ID      AAZ32021  standard; DNA; 3331 BP
vv
```

AC AAZ32021

DT 10-JAN-2000 (first entry)

DE Human METH1 related EST D86074.

KW Human; METH1; METH2; anti-angiogenic; metalloprotease thrombospondin;

KW angiogenesis inhibitor; abnormal wound healing; inflammation;

diabetic retinopathy; macula degeneration; haemangioma; detection;

[illegible]

XX  
05  
XX

PN MO9937000-AL  
XX

PD 29-JUL-1999.  
XX

PF 22-JAN-1999; 99WO-US0131313.  
XX

PR	23-JAN-1998;	98US-0072298.
PR	28-AUG-1998.	98US-0098539.

XX  
DA / TRITE / TRITE 2-ABT CDF T

PA (HAST/) HASTINGS G A.

[illegible]

XX	1000	ENDORA/END
XX	1000	ENDORA/END

XX

PT treating hyperproliferative disorders, cancers or autoimmune disorders

XX

[illegible]

CC metalloprotease thrombospondin (METH) proteins METH1 and METH2

angiogenesis both in vitro and in vivo. They can be used for treating

CC wound healing, inflammation, rheumatoid arthritis, psoriasis,

macula degeneration, haemangiomas, and arterial-venous malformations.

system, by activating or inhibiting the proliferation, differentiation,

immune deficiencies or disorders may be genetic, somatic, such as

CC cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or CC toxins), or infectious. They can also be used to treat inflammatory

used for detection and diagnosis. AAZ32002 to AAZ32080, and MAY49503 to

CC AY49511 represent sequences given in the exemplification of the present invention

Sequence 3331 BP: 762 A: 701 C: 1022 G: 846 T: 0 other:  
XX  
XX

```
alignment_scores: 381 50
Overlap: 1
Length: 597
```

Ratio:	1.408	Gaps:	25
Percent Tdonativ:	45.304	Percent Tdonativ:	25.639

US-09-668-119-3 x AAZ32021/rev ..

Align seg 1/1 to reverse of: AAZ32021 from: 1 to: 3331

---







Fri, Mar 1 09:12:02 2002

us-09-668-119-3.rng

---

Page 17











```

922 GGTACAGAGTGGTG ZAGCCCTTGCTGACGCCAAACAGTACTGTGAG 971
122 rProglInThrGlnLeuGlnGlnValAlaLeuGlnGlnGlnGln 139
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
972 CCAGGGCCAGCCAGACAGAGGCAAGATGCGACGCCAAGAAAGGACGAG 1021
139 InGlnGlnInPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 155
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1022 CAGATGGAGGTGGCCAGCAGATGGGCATGAACCTGACAGCAGCAGCC 1071
156 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 172
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1072 ACACCTGGCCCGACGACAGCATTATGCTCAGCCACCTACACAGAT 1121
172 nGlnPheGlnAlaValGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 189
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1122 CCAGCCCATTCACATCATTCAGCAACAGCAACAGATCCACCTCCAGCAG 1171
189 InGlnGln..... GlnHisLeuIleuLysLeuHisGlnAsnGlnGln 203
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1172 AACAGGTGGTATCCAGCAGCAGATGGCATCCACCC..... CAGCAG 1215
204 GlnIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 219
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1216 CAGTTCAGCAGCCGAGTCCACCTCCTTCACACAGCTTACACACCTTCA 1265
219 nLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 236
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1266 GTTGGCGCAGCAGACACAGCAGCAACACACACAGCAGCAGCAGCAGC 1315
236 In..... A..... AlaLeuGlnAlaGlnProGlnIleGlnProPro 248
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1316 AGCCGCAAGCCACCAACCTCCTACCTGCTCCACACAGCAGCAGCTTCA 1365
249 MetGlnProGlnProGlnProGlnProGlnProGlnProGlnProGln 265
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1366 ACTCAG..... CAGTCCACCTTCCACCTCAGTCC..... CA 1394
265 nGlnMetHisIleThrGlnHisIleGlnProProProGlnProGlnPro 282
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1395 GCACCAAGCCCAACCCGCTGCTCAG..... 1422
282 rOProValAlaGlnAsnGlnProSerGlnLeuProGlnProGlnThr 298
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1423 ..CCATGCTTCAGTCTTCACTGCTGCTCCTCAGCTGATGACGCCCT 1470
299 GlnProLeuValSerGlnAlaGlnLeuProGlnGlnMetLeuIleThr 315
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1471 AAGCCACCAATCC.....ATCCATC 1493
315 rGlnProProLeuLysPheValArgAlaProMetVal..... 327
1494 CAACCAACCTGTAGC/CCTTCAAGCCGCTGATTAGGGCCGCTAAGA 1543
328 .....ValGlnGlnProProValGlnProGlnValGlnGlnGln 341
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1544 TGTCACTGCTCCAGCAGCAACCA.....CCCATATCCTCTGCAAGTT 1587
342 ThrAlaValGlnThrAlaGlnAlaGlnMetValAlaProGlnValGln 358
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1588 GTAGGACTGACAGCAGGATACAGCCAGCAGCTTGGGGTTGGC 1637
358 nValSerGlnSerSerLeuProMetLeuSerSerProSerProGlnGln 375
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1638 ACAGCTGGCAGCTGCTGACT..... 1659
375 InValGlnThrProGlnSerMetProProProProGlnProSerProGln 391
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1660 .....ACTTCCCGGGGATGCGAGTACAGTCAAG...TCTGGTCA 1698
392 ProGlnProSerSerGlnProAsnSerAsnValSerSerGlnProAl 408
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1699 GCCCATTTGGCTCTGCTGACCTTATCCAGGCTCTGGTGCACTGCA 1748

```

```

408 aProSerProSerPheLeuProSerProSerProGlnProSerGln 424
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1749 GAGTGGCTTCCACATTTGGCCCTGGATGACCTTGTCTGCTGTCAG 1798
425 .....SerProValThr 428
1799 GAGACGACATGTGTAAAGGTGGGCTACACCTCTCAGCTGTGTA 1848
429 AlaArgThrProGlnAsnProSerProGlnProLeuAsnThr 445
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1849 GCCAGGTCTCTGCTGCTTATGACAGTCT.....GTGACTT 1889
445 rProValAsnProSerSerValMetSerProAlaGlnSerSerGlnAlaG 462
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1890 GCGGGTAAACCCAGCAGATG.....CTGTCAACCCAGCAGCTG 1930
462 InGlnGlnGlnIleuAspLysLeuGlnLeuSerLysIleGln 478
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1931 ACTGTAGGAGGAGAGATGATGCTCCACATTTGGCTCAATGCTCT 1980
479 ProLeuAlaArgMetIleAsnLysIleAspLysAsnGlnAspArgLys 495
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1981 GCCAAAGCATCTCCAGTACAGAACCCCAAAAGTATGACGAGAGAG 2030
495 AspLeu.....SerLysMetLysSerLeuAspIle..... 506
2031 CAGTCTTGGAGAAAAGCTGATCAAGTGTGCTATGCAATGCTAATATC 2080
507 .....LeuThrAspProSerLysArgCysProLeu 516
2081 CAAGCAGTGAATAGTACCTTGACC...CCGCCCCCTTCACTACGCT 2127
517 LysThrLeuGlnLysCysGlnIleAlaLeuGlnLysLeuAsnAspM 533
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2128 CCTACACTGACCATGCT.....CTAGCAAAAGGCTGACTCAAA 2168
533 eArgCysProLeuProHisArgPro 541
2169 AC.....CCCCAGCAGCCA 2182

seq_name: /cgn2_6/ptodata/2/1na/6A_COMB.seq:us-08-864-038A-1
seq_documentation_block:
: Sequence 1, Application US/08864038A
: Patent No. 6001592
: GENERAL INFORMATION:
: APPLICANT: KUNIO NAKASHIMA et al.
: TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
: TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
: TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
: TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBOD
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: 812-5 Hirano
: STREET: Isshinden
: CITY: Tsu-city
: STATE: Mie-prefecture
: COUNTRY: JAPAN
: ZIP: 514-01
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
: OPERATING SYSTEM: Microsoft Windows 95
: SOFTWARE: Word Perfect 6.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/864,038A
: FILING DATE: May 28, 1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 8-184459
: FILING DATE: 15-July-1996
: ATTORNEY/AGENT INFORMATION:

```











```

204 GlnIleGlnGlnGlnGlnGln.....LeuG1 213
    |||:|||||:|||||
1037 CAGCTGAGAGGCGGAGGAGGAGGAGCTGTGGGCTCTCTCTCTCTCA 988
    |||:|||||
213 nargIlealagIleuGlnLeuGlnGlnGln..... 224
    |||:|||||
987 CCGGCTCTACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 938
    |||:|||||
225 .....GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 238
    |||:|||||
937 ACCTCGGCGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 888
    |||:|||||
239 GluAlaGlnProIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 255
    |||:|||||
887 CATAT....CTCCGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 842
    |||:|||||
255 oProSerGlnAlaLeuProGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 272
    |||:|||||
841 ACCGAGTCTCTCA..... 829
    |||:|||||
272 iShsGlnProProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 288
    |||:|||||
828 ....ACTCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 796
    |||:|||||
289 ProSerGlnLeuProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 305
    |||:|||||
795 .....CTCTCTGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 767
    |||:|||||
305 agIleAlaLeuProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 322
    |||:|||||
766 TGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 717
    |||:|||||
322 alarGAlaProMetValValGlnGlnGlnGlnGlnGlnGlnGlnGln 338
    |||:|||||
716 CAGGCGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 670
    |||:|||||
339 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 355
    |||:|||||
669 .....GCAGCAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 626
    |||:|||||
355 ogIValGlnValSerGlnSerSerLeuProMetLeuSerSerSerP 372
    |||:|||||
625 ACCGAGTCTCTCA.....TCATCGAATAATCTA 579
    |||:|||||
372 roGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 387
    |||:|||||
578 AATCAATAATATGCTCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGG 529
    |||:|||||
388 ProSerProGlnProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 404
    |||:|||||
528 CCAAGTCTCACTCTCGAGAGCTCTCG..... 505
    |||:|||||
404 rSerGlnProAlaProSerProSerSerPheLeuProSerProSerP 421
    |||:|||||
504 .....CCAGATCCAAAGTCTCACT.....CCAGACCTGCTCTCTG 471
    |||:|||||
421 InProSerGlnSerProValThrAlaArgThrProGlnGlnGlnGln 437
    |||:|||||
470 CTCTGCGCTCTCTCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 421
    |||:|||||
438 ProSerProGlnProLeuAsnThrProValAsnProSerSerValMet 454
    |||:|||||
420 CCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 371
    |||:|||||
454 rProAlaGlnSerSerGlnAlaGlnGlnGlnGlnGlnGlnGlnGlnG 470
    |||:|||||
370 CCATTCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 321
    |||:|||||
471 LysGlnLeuSerLysTyrIleGlnProLeuArgArgMetIleAsnL 487
    |||:|||||
320 CACCGC.....CGCGCGCGCGCGCGC..... 300
    |||:|||||
487 eaSPlysAsnGlnAspArgLysLysAspLeuSerLysMetLysSerLe 504
    |||:|||||

```

```

239 .....CGTGGCTCTC 290
    |||:|||||
504 eaSPLeuLeuThrAspProSerLysArgCysProLeuLysThrLeuGln 520
    |||:|||||
289 TCACACACCGCGAGCTCTCCAGGCGCTCTCCGCGCTCTCTCTCTCT 246
    |||:|||||
521 LysCysGluIleAlaLeuGlnLysLeuLysAsnAspMetArgCysPro 537
    |||:|||||
245 .....CGTAAAGCTGCACTGCTCAT 223
    |||:|||||
537 uPro.....HisArgProAlaGlyCysHisArgPro 546
    |||:|||||
222 GCCGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 186
    |||:|||||
seq_name: /cgn2_6/ptodata/2/ina/6a_COMB.seq.us-08-864-038A-4
seq_documentation_block:
: Sequence 4, Application US/08864038A
: Patent No. 6001592
: GENERAL INFORMATION:
: APPLICANT: Kunio NAKASHIMA et al.
: TITLE OF INVENTION: NOVEL POLYPEPTIDE GENE CDNA, VECTOR
: TITLE OF INVENTION: CONTAINING SAID CDNA, HOST CELLS TRANSFORMED WITH SAID
: TITLE OF INVENTION: VECTOR, POLYPEPTIDE PRODUCED THEREBY, METHOD OF PRODUCING
: TITLE OF INVENTION: SAID POLYPEPTIDE, DNA ENCODING SAID POLYPEPTIDE AND ANTIBOD
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: 812-5 Hirano
: STREET: Ieshinden
: CITY: Tsu-city
: STATE: Mie-prefecture
: COUNTRY: JAPAN
: ZIP: 514-01
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
: COMPUTER: IBM compatible
: OPERATING SYSTEM: Microsoft Windows 95
: SOFTWARE: Word perfect 6.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/864,038A
: FILING DATE: May 28, 1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 8-184459
: FILING DATE: 15-July-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: C. Bruce Hamburg
: REGISTRATION NUMBER: 22,389
: REFERENCE/DOCKET NUMBER: F-5610
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212)986-2340
: TELEFAX: (212)953-7733
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3331
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: ORIGINAL SOURCE:
: ORGANISM: Pinctada fucata
: CELL TYPE: mantle epithelial cell
: FEATURE:
: NAME/KEY: CDS
: LOCATION: from 50 to 2263
: IDENTIFICATION METHOD: P (by similarity to some other pattern)
: US-08-864-038A-4

```

```

alignment_scores:
: Quality: 381.50
: Ratio: 1.408
: Length: 597
: Gaps: 25

```



Percent Similarity: 45.394 Percent Identity: 25.628

Alignment block:

US-09-668-119-3 x US-08-864-038A-4/rev ..

Align seg 1/1 to reverse of: US-08-864-038A-4 from: 1 to: 3331

```

50 AlaSerValSerAspPro.....MetAsnAlaLeuGI 60
   |||.....|
1680 GCGGCGAGCGGATCCACCTCCACGATTTGTCTTCCAAAGCCCTTCT 1631
   |||
60 nSerLeuThrGlyLeuPro.....AlaAlaGlyAlaA 71
   |||
1630 AAGTCTCTTCCTCCACCACTGATGACAGCGGCGGCTGCGAGCAG 1581
   |||
71 laGlylleGlyMetProProArGlyProGlyGlnSerLeuGlyGlyMet 87
   |||
1580 CTGCGGCGAGATCCCTCCACCAAGCGGCT..... 1552
   |||
88 GlySerPheGlyAlaMetGlyGlnProMetSerLeuSerGlyGlnProPr 104
   |||
1551 .....AGTCTCCAAAGTCCACCAAAACCTCTCC 1523
   |||
104 oProGlyThrSerGlyMetAlaProHisSerMetAlaValAlaSerThra 121
   |||
1522 ACCT.....GCACCGACGACGAGTATGTCACGGGCTAAAG 1488
   |||
121 larThrProGlnThrGlnLeuGlnLeuGlnAlaValAlaLeuGlnGln 137
   |||
1487 CTCCGCTCCGCTCCACCTCCAAAGTCCACCTCCGCTCCGACGACGCG 1458
   |||
137 nGlnGlnGlnGlnGlnPheGlnGlnGln..... 147
   |||
1437 AGAGCGACGAGTAGAKTCTCAGCTCCACCGCGCTCTCTCCGTTACC 1388
   |||
148 .....GlnAlaAlaLeuGlnGlnGln 154
   |||
1387 ACCTCCGTTACCAACGACGCTCCACCTCCACCTGGAAGCTGATGCAATG 1338
   |||
155 GlnGlnGlnGlnGlnGlnGlnPheGlnAlaGln..... 166
   |||
1337 CAGATGCGAAGACAGATGCTGAGCTTTCAATAGACTCGCAAGATCCCTT 1288
   |||
167 GlnSerAlaMetGlnGlnGlnPheGlnAlaValAla..... 178
   |||
1287 AAGGATCTCTCAAAATGCTCAATTGAGATTGTGTGCTTGTCTACTGCTGA 1238
   |||
178 ..... 178
   |||
1237 TGTCTTAGACAGAGGCTTTTGACAGAGGCTCCATCTCTGAGAGGTCTT 1188
   |||
178 ..... 178
   |||
1187 TAATGCTCTTAATGCTGCCCTTATACCACTATTGTCTAGAGCTGAA 1138
   |||
179 GlnGlnGlnGlnGlnLeuGlnGlnGlnGlnGlnGln..... 192
   |||
1137 GCAGATGACAGAGCGCTGACAGCGGACGACGACGACGCGCTCCGCC 1088
   |||
193 .....HisLeuLeuLeuHisHisGlnAsnGlnGln 203
   |||
1087 TCCCTCCCAACCACTCCACCTCTCTCCGCCACCAACGACGCGCTGCGGG 1038
   |||
204 GlnlleGlnGlnGlnGlnGln.....LeuGI 213
   |||
1037 CAGCTGACAGCGGCGACGACGACGCTGCGGCGCTCTCTTCTTCGGA 988
   |||
213 nArgTAlaGlnLeuGlnLeuGlnGlnGln..... 224
   |||
987 CCGGCTTACTCTCTGCTCCACCGTAGAAACCAACACGACGATCCCTCC 938
   |||
225 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnAlaLeu 238
   |||

```

```

937 ACCTCCGCGACAGCGGCGGACGACGACGACGCGGACGACGATCTCC 888
   |||
239 GlnAlaGlnProProIleGlnGlnProMetGlnGlnProGlnProPr 255
   |||
887 CATAT.....CTCCGAGGCGACCAAGACCTCCGAGGCTTCCAAAGTCCGC 842
   |||
255 oProSerGlnAlaLeuProGlnGlnLeuGlnGlnMetHisThrGlnH 272
   |||
841 ACCGAGTCTCCA..... 829
   |||
272 lshGlnProProProGlnProGlnGlnProProValAlaGlnAsnGln 288
   |||
828 .....AGTCCGCCACCGAGTCTCCAAATGTCACCTTCCG..... 796
   |||
289 ProSerGlnLeuProProGlnGlnGlnGlnGlnGlnGlnGlnGln 305
   |||
795 .....CTCTCTCAGCGGCGACA.....GCGGCTCCGC 767
   |||
305 aglnAlaLeuProGlyGlnMetLeuTyrThrGlnProProLeuLysPheV 322
   |||
766 TGCAGCAGCAGCTCTCCAAAGTCTACTGCGCTCTCTCCGACGCGCAG 717
   |||
322 aLArgAlaProMetValValGlnGlnProProValGlnProGlnGln 338
   |||
716 CAGCGGCTGGGCTGGCGGACGACGCTCCACCACTCCACCCCGCG... 670
   |||
339 GlnGlnGlnThrAlaValGlnThrAlaGlnAlaAlaGlnMetValAlaPr 355
   |||
669 .....GCACGACGCGGACGCTCCGACGACGACGACGACGACGCTCC 626
   |||
355 oglyValGlnValSerGlnSerSerLeuProMetLeuSerSerProSerP 372
   |||
625 ACCAGCTCCACGAGGCGAAGAGCTGCACCAAGA...TCATGGAATCTA 579
   |||
372 roGlyGlnGlnAlaGlnThrProGlnSerMetPro...ProProGln 387
   |||
578 AATCAATTAATGCTCTCCACGCGCCCAAGTCCGCCAAGACTCTCAAGTCCG 529
   |||
388 ProSerProGlnProGlyGlnProSerSerGlnProAsnSerAsnValSe 404
   |||
528 CCAGTCCACCTCCGACCTCCG..... 505
   |||
404 rSerGlyProAlaProSerProSerSerPheLeuProSerProSerProG 421
   |||
504 .....CCCAATCCAAATGCGAGT.....CCAGACCTGCTCTCG 471
   |||
421 InProSerGlnSerProValThrAlaArgThrProGlnAsnPheSerVal 437
   |||
470 CTCTGCGCGCTGCTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 421
   |||
438 ProSerProGlyProLeuAsnThrProValAsnProSerSerValMetSe 454
   |||
420 CCAGCAGCAGCAGCTCCGCGGCGACGCTCCCATTCACCGCTCATCTCCAGT 371
   |||
454 rProAlaGlySerSerGlnAlaGlnGlnGlnGlnGlnGlnGlnGlnGln 470
   |||
370 CCATTCATATGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 321
   |||
471 LysGlnLeuSerLysTyrIleGlnProLeuArgArgMetIleAsnLysII 487
   |||
320 CACCGC.....GCGCCCGCGCTCCG..... 300
   |||
487 eaSplysAsnGlnAspArgLysLysAspLeuSerLysMetLysSerLeu 504
   |||
299 .....CGTCCGCTCC 290
   |||
504 euAspIleLeuThrAspProSerLysArgCysProLeuLysThrLeuGln 520
   |||
289 TCCAGACCGCGCAGCTCCCAAGCGCTCCGCGCTCCGACAC..... 246
   |||
521 LysCysGluIleAlaLeuGlnLysLeuLysAsnAspMetArgCysProLe 537
   |||
245 .....CGTAAAGACCTGCACTGCTCAT 223

```



Downloaded from <http://ajph.org/> on November 10, 2014



Page 8

1. The first step is to identify the key components of the system. This involves understanding the hardware, software, and data involved in the process.

[illegible]







Fri Mar 1 09:12:02 2002

```

: TELEPHONE: (212) 278-0400
: TELEFAX: (212) 391-0525
: INFORMATION FOR SEQ ID NO: 20:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 32207 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: double
:   TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: DS-08-757-663A-20

```

alignment\_scores:

Quality:	314.00	Length:	343
Ratio:	1.602	Gaps:	9
Percent Similarity:	57.143	Percent Identity:	30.904

alignment\_block:  
US-09-668-119-3 x US-08-757-669A-20/rev

Align seg 1/1 to reverse of: US-08-757-669A-20 from: 1 to: 32207

```

7  AlahisSerLysser3erLyssaspmetGluserHisvalPheLeuLysAl 23
   ::::: ::: :::::
20489 AGCCACACAGCAGGAGCCACAGCAGCAGGAGCCACACAGCAGCAGGAGCCA 20440

```

```

23  alysthrargaspcljtyrleuserleuvalalargleullehisp 40
      :::::||||:
20439 CAGCAGCAGGAGC.....CACA 20423

```

40 heargspilehisamlyssyserglnalaservalser.aspprome 56  
|||:::||||:::||||||  
20422 GCAGCAGGAGCCACA3CAGCAGGAGCCACACAGCAGCAGGAGGCCACAGCAGC 20373

56 taaalalaleuGInsectleuthrGlyGlyProalaaGlyalaaGlyI 73  
::||| :: ::|||::| ||||| |||||::: :::  
20372 AGGAGCC....ACAGCAGCAGGACCCACAGCAGCGGGAGCCACAGCAG 20329

73 leglymetProProarglyProglyInserIæuglyglymetylser 89  
 ||| ||| ::| ||| ::| ||| ::|  
 20328 CGGGAGCCCCA...GCACGGGGAGCCACAGCAGCGGGAGGCACAGCA 20285

```

90 PheGlyAlaMetGly31nPrometSerLeuSerglyInProProG1 106
      |||||      ::      ::|||
20284 GCGGGAGCCACAGCA3CGGGAGCCACAGCAGCGGGAGCCACAGCGGG 20235

```

```

106 yThrSerGlyMetAlaProHisSerMetAlaValSerThraIaThr. 122
|:::| |:::| ::|||:::|
20234 AGCCACA.....GCAGCGGAGCCACGACGACG 20206
```

123 ProGlnThrGlnLeu;InLeuGlnGlnValAlaLeuGlnGlnGlnGln 139  
::: ||| ::: ||||| | |||||::: |  
20205 GATGAGCAGCAGCAG;ATTGAGCAGCAGCAGATGAGCAGCAGCATGA 20156

139 nGInGIInPheGInGIInGIInAlaAlaLeuGIInGIInGInG 156  
:::|||||  
20155 GCAGCAGCAGGATGAATACAGACCAGCATGAGCAGCAGCAGCACTTCCACCC 20106

**156** lnglnglnglnglnPheglAlaglngInSerAlamEtgIngln **172**

[illegible]

```

189 nglnglgnlhlsllellyslsluhsllslglnasnglgnlgnlleg 206
    :|||||||      :: : : : : : : : : : : : : : : : :

```

```

206 InGInGInGInGInLeuGInArgIleAlaInLeuGInLeuGIn 222
:::|||||::: : ||::: ||| :::|||

```

19961 AGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAG 19912

us-09-668-119-3.rni

Page 10

[illegible]



















```

153 GlnGlnGlnGlnGln.....GlnGlnGlnGlnGlnPheGlnAl 165
1046 CAACAACAACAACAACAACAACAACAACAACAACAACAACA 1095
165 aglnGlnSerAlaMetGlnGlnGlnPheGlnAlaValGlnGlnGln 182
1096 AGCACA.....CTACAACAACAACA 1115
182 GlnGlnLeu.....GlnGlnGlnGlnGlnGlnGlnHisLeuLe 195
1116 CACAACTACCAAGAAACAACAACAACAACAACAACAACAACA 1165
196 LysLeuHisHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 212
1166 .....CTACAACAACAACAACAACAACA 1188
212 uGlnArgIleAlaGlnLeuGlnLeu.....GlnG 222
1189 ACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1238
222 GlnGlnGlnGlnGln.....GlnGlnGlnGlnGlnGlnGln 233
1239 AACCAACAACAACAACAACAACAACAACAACAACAACAACA 1288
234 GlnGlnGlnAlaLeuGlnAlaGlnPropGlnGlnGlnGlnGln 250
1289 CAACAACAACAACAACAACAACAACAACAACAACAACAACA 1332
250 nGlnProGlnPro.....ProProSerGln.....A 259
1333 ACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1382
259 IalLeuProGlnGlnGlnGlnGlnGlnGlnHisHisHisHis 273
1383 TACTACCAACAACAACAACAACAACAACAACAACAACAACA 1432
274 .....GlnProProGlnProGln.....GlnProProValAl 285
1433 CTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1482
285 aglnHisnGlnProSerGln.....LeuProGlnGln 296
1483 AAGAAACAACAACAACAACAACAACAACAACAACAACAACA 1532
296 erglnThrGlnProLeuValSerGlnAlaGlnAlaLeu 308
1533 ACCAACAACAACAACAACAACAACAACAACAACAACAACA 1570

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-928-361B-1

seq_documentation_block:
Sequence 1, Application US/08928361B
Patent No. 6071518
GENERAL INFORMATION:
APPLICANT: Petersen, Carolyn
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
TITLE OF INVENTION: SPECIES INFECTIONS
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: PETERS, VERNY, JONES & BIRSA
STREET: 385 Sherman Avenue, Suite 6
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306-1840
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

```

```

APPLICATION NUMBER: US/08/928,361B
FILING DATE: 12-SEP-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,062
FILING DATE: 13-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Verny, Hana
REGISTRATION NUMBER: 30,518
REFERENCE/DOCKET NUMBER: 480.76-1(HV)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-324-1677
TELEFAX: 650-324-1678
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 7334 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-928-361B-1

```

```

alignment_scores:
Quality: 296.50 Length: 246
Ratio: 2.335 Gaps: 13
Percent Similarity: 51.626 Percent Identity: 41.057

```

```

Alignment_block:
US-09-668-119-3 x US-08-928-361B-1
Align seg 1/1 to: US-08-928-361B-1 from: 1 to: 7334

```

```

113 HisSerMetAlaValAlaSerThrAlaThrProGlnThrGlnLeuGlnLe 129
2564 CATACACTAATATGTGTGGAGTGAACACACACACACACACACACAC 2613
129 uGlnGlnAlaLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 145
2614 ACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 2663
146 .....GlnGlnGlnAlaAlaLeuGln 152
2664 TACTACAACACTACTACTACTACTACTACTACTACTACTACTACTACT 2713
153 GlnGlnGlnGlnGln.....GlnGlnGlnGlnGlnPheGlnAl 165
2714 CACAACAACAACAACAACAACAACAACAACAACAACAACA 2763
165 aglnGlnSerAlaMetGlnGlnGlnGlnPheGlnAlaValGlnGlnGln 182
2764 AGCACA.....CTACAACAACAACAACAACAACA 2783
182 GlnGlnLeu.....GlnGlnGlnGlnGlnGlnGlnHisLeuLe 195
2784 CACAACCTACCAAGAAACAACAACAACAACAACAACAACAACA 2833
196 LysLeuHisHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 212
2834 .....CTACAACAACAACAACAACAACAACA 2856
212 uGlnArgIleAlaGlnLeuGlnLeu.....GlnG 222
2857 ACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 2906
222 GlnGlnGlnGlnGln.....GlnGlnGlnGlnGlnGlnGln 233
2907 AACCAACAACAACAACAACAACAACAACAACAACAACAACA 2956
234 GlnGlnGlnAlaLeuGlnAlaGlnPropGlnGlnGlnGlnGln 250
2957 CAACAACAACAACAACAACAACAACAACAACAACAACAACA 3000

```



Fri Mar 1 09:12:02 2002

us-09-668-119-3.rni

Page 16

```

250 ngInProGlnPro..          ProProSerGln.....A 259
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
3001 ACACACCAACACCACTACTACACTACCAAGAAACCAACCAACTACTACTAC 3050
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
259  lateuProGlnGlnGlnGlnGlnMetHisHisThGlnHisHis..... 273
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
3051 TACTACCAACAACAACAACTACTACTACTACTACCAACAACAACAACACTA 3100
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
274  ..... GlnHroProGlnProGln.. GlnProProValAl 285
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
3101 CTACTACTACTACACACCAACAACAACAACACGACACCAACAACACTACC 3150
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
285  agInaGlnProSerGln..... LeuProProGlns 296
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
3151 AAGAAACCAACAACACCAACAACACTACTACTACTACTACTACCAACAAGAA 3200
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
296  erGlnThGlnProLeuValSerGlnAlaGlnAlaLeu 308
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
3201 ACCAACACAACTACTACTACTGCGACACAAACAACACTACTACTA 3238

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 28, 2002, 19:34:08 ; Search time 20.75 Seconds  
(without alignments)  
2125.545 Million cell updates/sec

Title: US-09-668-119-3

Sequence: 3010  
1 MKRAGVASHKSSKDMESHVF.....WPTSAHLSSTIPCTAHSFQP 579

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 68:\*

1: p1r1: \*  
2: p1r2: \*  
3: p1r3: \*  
4: p1r4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	447	14.9	796	2 T20393	hypothetical prote
2	439	14.6	4957	2 T03455	ALR protein - huma
3	439	14.6	5262	2 T03454	ALR protein - huma
4	399	13.3	1893	2 A56158	eye development pr
5	397	13.2	2251	2 T24490	hypothetical prote
6	384	12.8	966	2 S25365	CYC8 protein - yea
7	383.5	12.7	1012	2 I53172	RAE-28 - mouse
8	383	12.7	708	2 A53185	G-box-binding fact
9	378	12.6	1596	2 A33106	neurogenic locus m
10	377	12.5	1307	2 T25563	hypothetical prote
11	377	12.5	1655	2 T13998	gene mastermind pr
12	372	12.4	1234	2 T30160	hypothetical prote
13	365	12.1	1589	2 T13606	hypothetical prote
14	351	11.7	1390	2 T14004	trif protein - sil
15	345.5	11.5	1145	2 T18235	transcription acti
16	345.5	11.5	1457	2 T14577	protein kinase yak
17	342	11.4	947	2 T23107	hypothetical prote
18	337	11.2	1074	2 T24877	hypothetical prote
19	335	11.1	374	2 T05923	glutinin low molec
20	333.5	11.1	613	2 S27770	hypothetical prote
21	330.5	11.0	1076	2 T18236	omega secalin prec
22	330	11.0	1076	2 T24887	hypothetical prote
23	327.5	10.9	1008	2 S19033	transcription acti
24	327.5	10.9	2038	2 A43742	female sterile hom
25	326.5	10.8	905	1 R6B355	regulatory protein
26	325.5	10.8	357	2 S18235	omega secalin prec
27	323.5	10.7	359	2 T06982	glutinin low molec
28	323	10.7	981	2 T16060	hypothetical prote
29	322.5	10.7	684	2 A56154	Abl substrate ena

30	320.5	10.6	815	2 B30843	glutinin high mole
31	320	10.6	930	2 T08588	hypothetical prote
32	318	10.6	1180	2 S69205	stripe a/b protein
33	314	10.4	139	2 A26892	Mopa box protein -
34	313.5	10.4	815	2 JN0689	glutinin, high-mol
35	313.5	10.4	830	2 S15720	glutinin high mole
36	313	10.4	534	2 G86385	hypothetical prote
37	312.5	10.4	3190	2 T13828	CREB-binding prote
38	310	10.3	255	2 A60637	merozoite antigen
39	309.5	10.3	1761	2 T13675	hypothetical prote
40	308	10.2	356	2 S01992	glutinin low molec
41	307.5	10.2	2578	2 A56922	transcription fact
42	304.5	10.1	2783	1 A41948	alpha-fetoprotein
43	301.5	10.0	1062	2 G86325	hypothetical prote
44	300.5	10.0	788	2 T25061	hypothetical prote
45	298.5	9.9	838	1 EEWTHW	glutinin, high mol

## ALIGNMENTS

RESULT	1
T20393	hypothetical protein DY3.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans	
C:Date: 15-Oct-1999	#sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T20393	
R:Lennard, N.	
A:Submitted to the EMBL Data Library, June 1997	
A:Reference number: Z19266	
A:Accession: T20393	
A>Status: preliminary; translated from GB/EMBL/DBJ	
A:Molecule type: DNA	
A:Residues: 1-796 <Full>	
A:Cross-references: EMBL:Z96047; PTDN:CAB09414.1; GSPDB:GN00019; CESP:DY3.5	
A:Experimental source: clone DY3	
C:Genetics:	
A:Gene: CESP:DY3.5	
A:Map position: 1	
A:Introns: 56/1; 90/1; 204/1; 345/3; 379/3; 406/3; 649/3; 688/3; 727/3; 759/1	
Query Match	14.9%; Score 447; DB 2; Length 796;
Best Local Similarity	32.0%; Pred. No. 3.9e-13;
Matches 163; Conservative 47; Mismatches 171; Indels 128; Gaps 24;	
QY	50 ASVSDPMNALQSLTGAPAGAGICMPRPGQSLGKMS-----FGAMQPMLSGQP 103
DB	303 AKYSQVLDILKRAVGIQLPGE--MSMOPNTATNAGYQTSQQQGVANQVDTYGGSSQP 360
QY	104 PP-GTSGMAPHSMAYVSTATPTQQLQQLVLAQQQQQQQPPQQQQAALQQQQQQQQQ 162
DB	361 PRTQYTSYGVQNVV-----QPQPQQA-----QQQQLQQQQQLAABQKQKQQQQQQ 406
QY	163 PQAQSGAMQQAQVAVVQQQQLQQLQQLQQLIKLHNQQLQQLQQLQQLQQLQQLQQLQ 222
DB	407 IADQ-----QLAEQIAAQQAQREARLAQAQAQAQAQAQLQQLIRQQQLQ 455
QY	223 QQQQQQQQQQ-----QQQQQALQAPRIQQRPMQQRPPPSQALPQQLQQLHNQHNPP--- 276
DB	456 QQQQQQQQQQYGGKQPMRYVQVQAPVQVQ-----QIAQPPQPPQAPVQ-----QPMNPLPAPAPITQ 510
QY	277 -PQPPQPPVQAQNPQSLPPOSQTOP-----LVSAQAQLPQ-----MLTQPLKFFV 322
DB	511 APAPQAP--APMQPQPPQVYVQPTTKNNQTAALQGVNTYSQSKRPQVYTYTTPGSPQV 568
QY	323 RAV-----MVVQPPVQVQVQVQVQVQTAQAQAQVAVGVQVQSSQLPMLSSPS 371
DB	569 QIPHTSTYGNHEDSVIEEPHGSVLSQQ-----VVP-----PAPVQPV 608
QY	372 PCQGVQVTPQSMPPPPQPPQPPQPS--QPNSSVSSGAPRPSRSLPSPSPQSPQSPVTA 429
DB	609 PPMSTVATVTRGPPPPV-----RPASVIAPISSSV--PKSQPR--PSTTRPPTRATIR 657







OY 441 ----GRLTPVNPSSVSPAGSS-----QAEEQYLKDLKOLSKYTEP 479  
 DB 4096 SKGLGPWPDPRLAETQKPEQSSLVPHGLDQVNGVPEASQLSIKQEP 4144

## RESULT 4

A56158

eye development protein canoe - fruit fly (Drosophila melanogaster)

Species: Drosophila melanogaster

Date: 03-Oct-1995 #sequence\_revision 03-Oct-1995 #text\_change 21-Jul-2000

Accession: A56158

C:Species: H. Nihomatsu, I.; Kondo, S.; Ueda, R.; Togashi, S.; Hirata, K.; Ikegami, Y

Genes Dev. 9, 612-625, 1995

A&gt;Title: canoe encodes a novel protein containing a GIGF/DHR motif and functions with NC

Accession: A56158

Status: preliminary; not compared with conceptual translation

Molecule type: mRNA

Residues: 1-1893 &lt;MTV&gt;

C:References: GB:D49534; NID:g705386; PIDN:BA08478.1; PID:g705387

C:Genetics:

A:Gene: cno

A:Cross-references: FlyBase:FBgn0000340

C:Superfamily: GIGF domain homology

F|848-924/Domain: GIGF domain homology &lt;GIG2&gt;

## Query Match

13.3%; Score 399; DB 2; Length 1893;

Best Local Similarity 27.4%; Pred. No. 1.le-10;

Matches 182; Conservative 58; Mismatches 196; Indels 228; Gaps 30;

OY 7 AHSKSKMESHVFLKATKREYLSIVARLLHFRDINKKSQASVSDPMALSLTNGP 66  
 DB 1010 AALPMSKSVPA---LHHHTSGTISLA-----NSKSRST---BSLHNNTSGM 1050  
 OY 67 A--AGAAGTGM--PRPGSGSLGSGMS-----FGMGCPMSLSGCPPTGSGM 110  
 DB 1051 GGIGGAAGGMLGPRNGSONNANGNGNENGFGYONLSVYRAQNSQPTLNERPIAHA 1110  
 OY 111 AHSMAVNVSTATPTGTOLQLOVALQDQDQDQDQDQDQDQDQDQDQDQDQDQDQ 154  
 DB 1111 AMNAVNGSSPLARQDQ 1170  
 OY 155 QQQQQQQQFQAQASMQQF-----QAVVQDQDQDQDQDQDQDQDQDQDQDQDQDQ 197  
 DB 1171 QQQQQQQQF---QHSLLQDQDQFALSSGNLNGQDQDQDQDQDQDQDQDQDQDQDQ 1227  
 OY 198 -----HNNQDQ---IQQQQQQDQDQ----- 214  
 DB 1228 QMMAPSMNISMNHNHQQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQ 1287  
 OY 215 -----TAQLQLQDQ 253  
 DB 1288 YRSQDLDPSTLYEQDQ 1345  
 OY 254 PRPSQA-LPQDQ 283  
 DB 1346 VPRSTAPRQDQ 1405  
 OY 284 --VAQNSQLRPOSOTOP-----LVSQAQLGQMLTYQRP 318  
 DB 1406 SYVASNGQNNRPLHSGSNPWEREREKELMKREHINQWREQDQDQDQDQDQDQDQDQ 1465  
 OY 319 LK-----PYARPVVQDQ 371  
 DB 1466 LKTLLEDFERRAQELQEQ---EEDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQ 1520  
 OY 372 PQGQVQGTQSMRPPRPPRPPRPPRPPRPPRPPRPPRPPRPPRPPRPPRPPRPPR 427  
 DB 1521 ---YQTEIKLAEMV-----DSNLSVDVSP--POP---PAPPAQLPSLNTQDPK 1561  
 OY 428 TARTPQNSVSPGDLNTPVNPSSVSPAGSSQAEEQYLKDLKOLSKYTEIPRLRMINK 487

DB 1562 SILKHNRYSEGQVPGSGAPSPSKQSKASFAEMRHLEHNPISNLAKELNQL-TMLDKD 1620  
 OY 488 DKNE 491  
 DB 1621 NNNE 1624

## RESULT 5

T24490

hypothetical protein T05A10.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C&gt;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000

C:Accession: T24490

R:Sturton, J. submitted to the EMBL Data Library, November 1995

A:Reference number: Z19898

A:Accession: T24490

A&gt;Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-2251 &lt;WII&gt;

A:Cross-references: EMBL:Z68108; PIDN:CAA92133.1; GSPDB:GN000028; CESP:T05A10.1

A:Experimental source: clone T05A10

C:Genetics:

A:Gene: CESP:T05A10.1

A:Map position: X

## Query Match

13.2%; Score 397; DB 2; Length 2251;

Best Local Similarity 29.5%; Pred. No. 1.6e-10;

Matches 177; Conservative 52; Mismatches 194; Indels 176; Gaps 24;

OY 121 ATPGTQLQLQVALQDQ 168  
 DB 151 ARHQAEQQAQQAQQAQQAQQAQQAQQAQQAQQAQQAQQAQQAQQAQQAQQAQQAQ 209  
 OY 169 AMQDQFQAVVQDQ 222  
 DB 210 LQDQDQLDQDQDQLDQDQDQLDQDQDQLDQDQDQLDQDQDQLDQDQDQLDQDQD 269  
 OY 223 QQQQQQQQDQ 282  
 DB 270 QD 322  
 OY 283 PVANQPSQLRPOSQTOPLVS-----QAQLRQSMILTYQRPKYARPVVQDQ 330  
 DB 323 --AQNNAQQRPSVASTPRLSLTPQLNDLTPQWQDQ-LQDQDQLDQDQDQDQDQDQ 378  
 OY 331 PRVQDQVQDQDQ-----TAVQTAQAQAV--ARGVVQSSSLRPMSSPFGQVQYT 378  
 DB 379 AOLAQQAQDQ 438  
 OY 379 POSMPPRPQ-----PSPQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQ 429  
 DB 439 LQNGAREAQHROLLISTTRAPRGITMGTPTIGLARREQDQDQDQDQDQDQDQDQDQ 498  
 OY 430 RTDQNSVSPGRLNRPVNDQ-----VMSPPAGSSQAEEQYLKDLKOLSKYTEIPRL 484  
 DB 499 PRMKNQSNPMSNPSTSTASATSHQILAPLSLKEPQ-----SSSKAASGNSMSDHI 555  
 OY 485 NK-IDKNE-----DRKK----- 495  
 DB 556 SRIISENEVLLQGDVYLRKKRPNYHRIQAQSSVVDHDSNCGSTRTSPGPKDSMILQAA 615  
 OY 496 -----DLSKKMSLIDILT-----DPSKR-----CPL----- 516  
 DB 616 SQSLFELSSGSKNHMGSLTSGQRLRLPQAHNDNPNYPRCICLYKLTTPRNEAGLQAHNVY 675  
 OY 517 --KTLQKCEIALEKLNKDMRCPLPHR--PRCHRPNSSTYASRSWMPSTSAHLSSTIP 571  
 DB 676 KKKELEKAQIAOE-----GNPHSLKRRHTHODATLAMHS-----PLAAHTPSPMP 721







[illegible]

```

RESULT      8
A:53185
G-box-binding factor - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 24-Sep-1999
C:Accession: A53185
C:Schmitzler, G.R.; Fischer, W.H.; Firtel, R.A.
Genes Dev. 8, 502-514, 1994
A:Title: Cloning and characterization of the G-box binding factor, an essential component
A:Reference number: A53185; MUID:94170994
A:Accession: A53185
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-708 <SC5>
A:Cross-references: GB:I29075; NID:g456561; PIDN:AAA21021.1; PID:g456562
C:Superfamily: G-box binding factor
C:Keywords: DNA binding; transcription factor; zinc finger

```

```

Query Match 12.7% Score 383; DB 2; Length 708;
Best local Similarity 24.4%; Pred. No. 2.3e-10;
Matches 170; Conservative 57; Mismatches 200; Indels 270; Gaps 23

OY 44 HNKKSQASVSDPPNNAOLSLTGCFRAGAAIGMPRG-----PGGS-LGG 86
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 6 HNGGNSSSSSSSSSPSQTGGSDLSNLSALRPLPSFTGAOMNMRPLFPTSSLGG 65

OY 87 MGSFGAGMRMS--LGGRRPRTSGMARSHMAVUS-----PLRTQTLQLOLUALQOQ 138
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 66 SSNTPFLLRPSSIMSSNVFPSSHDGQYRDMRNVDQYQINPNMRPNINUYQLMFMQQA 125

OY 139 QQQQFQQQQQAAALQQQQQQQQQQQQQFQAQDSFAMQDQFQAVVQQQQQQQQQQQQNHLIK 198
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 126 QQQNRRQQNQ--QQQNNQQQQQQRRNNQQMQQQQQNNQQMQQQQQQQQQQQQQQQQQ 182

OY 199 HNNQQQIQ--QQQQQLQQLAQQLQQLQQLQQLQQLQQLQQLQQLQQLQQLQQLQQL 255

```

[illegible]

RESULT 9  
A33106  
neurogenic locus mam protein - fruit fly (Drosophila melanogaster)  
N:Alternate names: mastermind protein  
C:Species: Drosophila melanogaster  
C:Date: 07-Sep-1990 #sequence\_revision 07-Sep-1990 #text\_change 21-Jul-2000  
C:Accession: A36391; A33106; S1514  
R:Smoller, D.; Friedel, C.; Schmid, A.; Bettler, D.; Lam, L.; Yedvobnick, B.  
Genes Dev. 4, 1688-1700, 1990  
A:Title: The Drosophila neurogenic locus mastermind encodes a nuclear protein unusual  
A:Reference number: A36391; MUID:91065516  
A:Accession: A36391  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1596 <SMO>  
A:Cross-references: GR:X54251; NTD:g8203; PIDN:CAA38152.1; PID:g8204  
A:Note: stratin Canton S  
C:genetics:  
A:gene: FlyBase:mam  
A:Cross-references: FlyBase:FBgn0002643

	Query Match	12.6%;	Score 378;	DB 2;	Length 1596;
	Best Local Similarity	30.2%;	Pred. No. 8,1e-10;		
	Matches 157; Conservative		Mismatches 143;	Indels 190;	Gaps 24;
QY	46 KKSQASVSDPMNALIQ-SLTGCPAAGAACIGMPRGCSOSTLGGSGFPGMGOPMSLSGP-	103			
Db	537 KVEPPANODILINLNKSEBGLGHGGFGSLGDINDGMKMRG-----GNPNOGGFPFN	589			
QY	104 -PEGTSGMADH-----SMAVVSTATPOTOLDLOVALQ-----	135			
Db	590 GPNGGTGCAGNACNGNGNSGIMSEHPLAQTLLKMAEHOHQHKNAMGGMGGPRPRPHKN	649			
QY	136 --GGG--GGGPFQGGGGAALQQ-----QQGGGQQGGGF	163			
Db	650 PQQGQQGQQGQQGQQQAQQQHGGMMGGGQPGRYNDYGSGFFNDEGLQPNQGQQGQQQAQQQ	709			



```

OY 164 QAOGSAM-----OOFOAVVO-----OOOOLQOOOQ 190
Db 710 OROOHNLPRFHOXKGRPGAGMNVQNFLLIKQELFSSQNDPDLKLOOQOAMQOQO 769
OY 191 OOHILKLNHNOQOILQ-----OOOOLQRTAQLOLOOQOQOQOQOQOQOQOQOQOQO 241
Db 770 QO-----HHOQOQYKMGVGVFNKQOQOQOQVPRQOLOOQOQOQOQOQOQOQOQOQOQO 824
OY 242 -----PP-----TOQP-----PMQOPRPPSOALPQOLOQMHNT-----QHQP----- 275
Db 825 PMAANFLNCPFRKIGPNSNDQPRGMLAQOQOQOQ-----GAGPQOQOQOQOQOQOQOQOQOQO 881
OY 276 -----PPOQOPPVVQOQPSQLPPOSOQOPLVSO-----AQAIPQOMLYTOPPLKFRV 323
Db 882 GGNTPMAPOOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 939
OY 324 APVAVQOPPVV--QIQQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 381
Db 940 VSAVADQOQVFPSCQAQOQOQO-----OOOQOQO-----GTNGPRPQOQOQOQOQOQO 981
OY 382 MP-----PPOQOPQOQPSQOQPSNVSSGPAPS 410
Db 982 GNAGGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGVGV 1021

```

## RESULT 10

hypothetical protein C24A8.3 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
 C:Accession: T25563  
 R:Neilson, J.; Galltling, S.  
 Submitted to the EMBL Data Library, December 1996  
 A:Description: The sequence of C. elegans cosmid C24A8.  
 A:Reference number: 220051  
 A:Accession: T25563  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1307 <NLS>  
 A:Cross-references: EMBL:U80845; PDB:1ABJ7980.1; GSPDB:GN00028; CESP:C24A8.3  
 A:Experimental source: strain Bristol NZ; clone C24A8  
 A:Genetics:  
 A:Gene: CESP:C24A8.3  
 A:Map position: X  
 A:Introns: 42/3; 116/2; 165/1; 200/3; 233/3; 275/1; 318/1; 358/3; 397/3; 590/3; 606/3; 6

Query Match 12.5%; Score 377; DB 2; Length 1307;  
 Best Local Similarity 23.4%; Pred. No. 7.5e-10;  
 Matches 169; Conservative 73; Mismatches 200; Indels 280; Gaps 31;

```

OY 56 MNALGLTSGPA-----AGAAGIGMP-----RGPGSLGKMGSGCA 92
Db 591 MOOQNNMNPBPSHP IOMOQGGQSTINRPSOPRHOQOQGGPSMNPBPSQOQHOGGPRQL 650
OY 93 MGPMSLSQPRPGSGM-----APHSMAVUSTATPOTQLOQOYA----- 133
Db 651 MNPBSSOMKVPORG--GMSRPPSOPRNYIQOGRPSK-----NTPRQOQMMQOIQOQNNNM 704
OY 134 -----LQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 163
Db 705 NNGSPHQOQPLQHQIMQOANRMMNMRPOQOIRIQOQOQOQOQOQOQOQOQOQOQOQOQOQO 764
OY 164 QAOGSAMQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 201
Db 766 RVADQVADQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 820
OY 202 QOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 250
Db 821 QPQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 880
OY 251 QPQPRP--SQALPQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 279

```

```

Db 881 QMOQDPMSQOPLQESDPNHLQHNQTPPMQOQOMQPVVTDNFEEDNDQITIRNGENHAI 940
OY 280 -----OOPEVAQNDPSOL-----PQSO-----TQPLVSQ 304
Db 941 NGMOYKLVPRQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 1000
OY 305 AQALPQOMLYTOPPLKFRVAVV-----VOOPRQOYQOQOQOQOQOQOQOQOQOQOQOQO 349
Db 1001 QONPIQO--HQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 1056
OY 350 A-----QMVAPGVQOQSSILPM-----LSSPFGQOQO 377
Db 1057 APRPRYRGGKNPAQOEPPIQOQTNQEDLNNSGILPQKQGRKVMADSLTNTFHQOQO 1116
OY 378 TPQSMPPRPPQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 428
Db 1117 APNGLP-----PNNAGPMIQOQKNITAGGYPRQOQOQOQOQOQOQOQOQOQOQOQOQO 480
OY 429 ARTPQNEVSPSPCLNTPVNS-----SVNSPAGSSQAEQOQYLDKIKQSKYI--EPL 480
Db 1171 AKSSPDLSPQSTSKORPPRANKADMTVIRDAIVSHVKGSRKMTPEYIE-----YVCBEL 1223
OY 481 RMINKIKDNEDRKD--LSKKKSLDILTPDSKRCPLTKQCEIALEKIKNDRCPL 538
Db 1224 NRMLELEK-ADMKTDRTYISVKRLALIQE-----EKLIITIPP-----RIQMP 1267
OY 539 HR 540
Db 1268 AR 1269

```

## RESULT 11

gene mastermind protein - fruit fly (Drosophila virilis)  
 C:Species: Drosophila virilis  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 17-Nov-2000  
 C:Accession: T13998  
 R:Newfeld, S.J.; Tachida, H.; Yedvobnick, B.  
 J. Mol. Evol. 38, 637-641, 1994  
 A:Title: Drive-selection equilibrium: homopolymer evolution in the Drosophila gene ma  
 A:Reference number: 217850; MUID:94365848  
 A:Accession: T13998  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1655 <NEW>  
 A:Cross-references: EMBL:M92914; NID:g157833; PID:g157834; PDB:1AC37201.1  
 A:Genetics:  
 A:Cross-references: FlyBase:FBgn0013119

Query Match 12.5%; Score 377; DB 2; Length 1655;  
 Best Local Similarity 25.2%; Pred. No. 9.3e-10;  
 Matches 162; Conservative 30; Mismatches 159; Indels 292; Gaps 19;

```

OY 64 GCPAAGAAGMPRRPG-----GSLGKGSF----- 90
Db 610 GCPYAGGGGG-----NGRGIMSEHSILAQTLKQMAEQOHNKMSMGSGGPNVPRHGMQO 666
OY 91 -----GAMGQPMLSGQPRP-----GTSGMAPHSMAVUSTATPOTQ-- 126
Db 667 QPQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 726
OY 127 -----LQLOQVALQOQO-----QOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 155
Db 727 HLPQFHQKARGGGPGVNVQOQNFLLIKQELFYSSPNFDLKNLQOQOQOQOQOQOQOQOQOQO 783
OY 156 QOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 196
Db 784 QOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 843
OY 197 LHNQ-----QOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 214

```



```

Db 844 FSNMANANFLNCPRGPGQGNAPGNMPPQ0000QPPRGPGQSNPNAPGGMAN 903
QY 215 IADLQ0000000000000000ALEA----- 240
Db 904 ATQ00000000000000000000ATTTTLQMKQTQDLHISQGGSHGIVSAGQLH 963
QY 241 -----OPRIGQPPMOQPPPPSOPALPQOLQOMHNTQH-----H 273
Db 964 LSSDKMSNVSVAACGVFFSSQ00AA00000000PQGNAPNQQ0000QPHGNGNAGNGG 1023
QY 274 QPPPGQPP-----PVAQNPQSLPPQSQTQPLVSAQALPGLMYLTQPL 319
Db 1024 PNGPQ00QPNQNMNNNSVSDGFLSLSQSMNFTQ00000000AAAAA0000QA----- 1079
QY 320 KFRAMVWVQPPVQVQ00Q0TAVQTACAAQNVARGVVSQSLPMLSPSPG----- 373
Db 1080 -----AAAQ0000QVPPNMQROTQACAAAAA-----AAQA0AAANANGPGGVPLM 1131
QY 374 -QGVTPQSMPPPPQPPQPPSQPNNSVSGPAPSPSFLPSPQPPSQPYTARTP 432
Db 1132 Q000Q0P-----GCVVVGAGSGNASVG-----VPVSAGSP 1161
QY 433 QNFS-----VPSGCLNTPPNPSVSPAGSSQAE 462
Db 1163 NNGAMNQLGPGMGMPQMOMGGPGVPIPMQMNPGAPNAQ 1204

RESULT 12
T30160
hypothetical protein C37A2.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T30160
R:Je, T.T.; Kemp, K.; Scheet, P.
submitted to the EMBL Data Library, April 1997
A:Description: The sequence of C. elegans cosmid C37A2.
A:Reference number: Z20746
A:Accession: T30160
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1234 <LEFT>
A:Cross-references: EMBL:U97194; PIDN:AAB52447.1; GSPDB:GN00019; CESP:C37A2.2
A:Experimental source: strain Bristol N2; clone C37A2
C:Genetics:
A:Gene: CESP:C37A2.2
A:Map position: 1
A:Intons: 46/1; 124/1; 159/1; 231/1; 508/3; 740/3; 891/1; 931/1; 975/1; 1014/3; 1083/3

Query Match 12.4%; Score 372; DB 2; Length 1234;
Best Local Similarity 25.4%; Pred. No. 1.2e-09;
Matches 170; Conservative 50; Mismatches 169; Indels 280; Gaps 29;

QY 71 AGTGMPPRGPG---QSL---GMSGFGAMGQPMSLSGQPPGTSGMAPHMAVSTAN 122
Db 403 SSGSGPPTVVGTPQPPQITPQPSLCPMGLPPTA-----PFGSQPMNP----- 447
QY 123 PQTQLDLOVVALDQ000000FQ0000A-----LQ00000000QF---QAQ0 167
Db 448 -----QQRITQ000000AAPSASNSPLLNLNSNQPPQ00YMPGPSA0G 491
QY 168 SAMQ0GFQAVVQ0000LQ000000QHLIKLHNQ0QI000000QLRIAGLQDQ00000 227
Db 492 LSMQ0--IAAIDQ0000-----HQYIQ0RIIDQ0000-----QAMMQ0 526
QY 228 Q00000000A-----LEA0PP-----IQ0PPW----- 249
Db 527 Q0QVQ0000QAAPPTTMRGHRGPFRTNMGAPRGTEGRIPRYAMGQPRMYHAGPQGM 586
QY 250 -----Q0P0PPSA---LPQ0LQ0MHTQHN0RPP0P0PVAQ0NP0SLP- 294
Db 587 IQRMNSTRGNACQFRPPQ0SQ0PPIPQ000---QPPAPQ0LQ0RPVQDTSVAVERPK 642

```

```

QY 295 -----Q0TQPLVSAQALPQOMLYTQPLKFRVAP-----MVVQ0P- 331
Db 643 KKKRTTKQKFAAAALAEQ000000Q0MAVYVQ000000Q0BQM000000G00Q0G 702
QY 332 -----PVQPVQ00Q0TAVQTAQA---QNVAPGVVSQS 362
Db 703 YPGQGYPMPPPGCAFPPGYLPGGAQPPQ0QIMQ0000LQ00RMA0PQ0000Q0GCGP 762
QY 363 SLPM-----LSSPSGQ0VQTPQSM-----PPP-----P 386
Db 763 Q0PMQ0MPQ0Q0LQVPPYPPGNNSTPDVAVN000NP1PGATMQRHLRSGEAPPPVSSGVHP 822
QY 387 QPSPPQ0PSSQPNNSVSGPAPSPSFLPSPQPPSQPYTARTPQNFSPSPGPIINTP 446
Db 823 Q000Q0GSHRDSASVSG-----SHTPGQ0GPGS0S0PTA-----VP-PGQ0N- 867
QY 447 VNPSSVSPAGSSQAEQ0YLDKQLK0LSKYIEPLRM-----INKID----- 488
Db 868 -----NPGSGSDIGEKAIVDOLNLS---EP1ADLDGLDGLDLEIPMDVQDQTPS 917
QY 489 -KNEDKKDLKMSLIDLTPDSKRCPLKTLQKCE1ALEK1KNDMRCPLP----- 538
Db 918 TSGERNERSDRLAISTELVDEVAAGRAGAFNAELAIIGRRGSVPIPLPASPGRNL 977
QY 539 ---HRRCH 544
Db 978 SHEHNMRAH 986

```

```

RESULT 13
T31606
hypothetical protein 87B1.5 - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T31606; S23632
R:Murphy, L.; Harris, D.; Bartell, B.
submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: Z17668
A:Accession: T31606
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1589 <MUR>
A:Cross-references: EMBL:Z96269; NID:el355202; PID:el251078; PIDN:CAB10975.1
R:Decamillis, M.; Cheng, N.; Plerre, D.; Brock, H.W.
Genes Dev. 6, 223-232, 1992
A:Title: The polyhomeotic gene of Drosophila encodes a chromatin protein that shares
A:Reference number: S23632; MUID:92146957
A:Accession: S23632
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1010; V, 1012-1192, 'L', 1194-1274, 'T', 1276-1589 <DEC>
A:Cross-references: EMBL:X63672; NID:g11056; PIDN:CA445211.1; PID:g11057
C:Genetics:
A:Gene: FlyBase:ph-p
A:Cross-references: FlyBase:FBgn0004861; FlyBase:FBgn0004860
A:Intons: 12/2; 595/1; 745/2; 1340/1
C:Superfamily: SAM homology
C:Keywords: DNA binding; nucleus
F:74-80; 247-285; 411-450; 494-650; 727-737; 775-955; 1032-1061/Region: glutamine-rich
F:1510-1576/Domain: SAM homology <SAM>

```

```

Query Match 12.1%; Score 365; DB 2; Length 1589;
Best Local Similarity 29.7%; Pred. No. 3e-09;
Matches 158; Conservative 57; Mismatches 177; Indels 140; Gaps 24;

QY 7 ASKSKSMDESHVPLKATRDXY-----LSVA---RLIHFRDHNHKSQA 50
Db 674 ASSVSTQTAQ0NSLKKAKRNQ0PYRALATLKTEIGQVAGQNKVVCHLTTVQ00Q0AT 733
QY 51 SVSDPMNALQSLTGGPAGAGAGIGMPRPGQSLGMSGFGAMGQPMSLSGQPPGTSGM 110

```



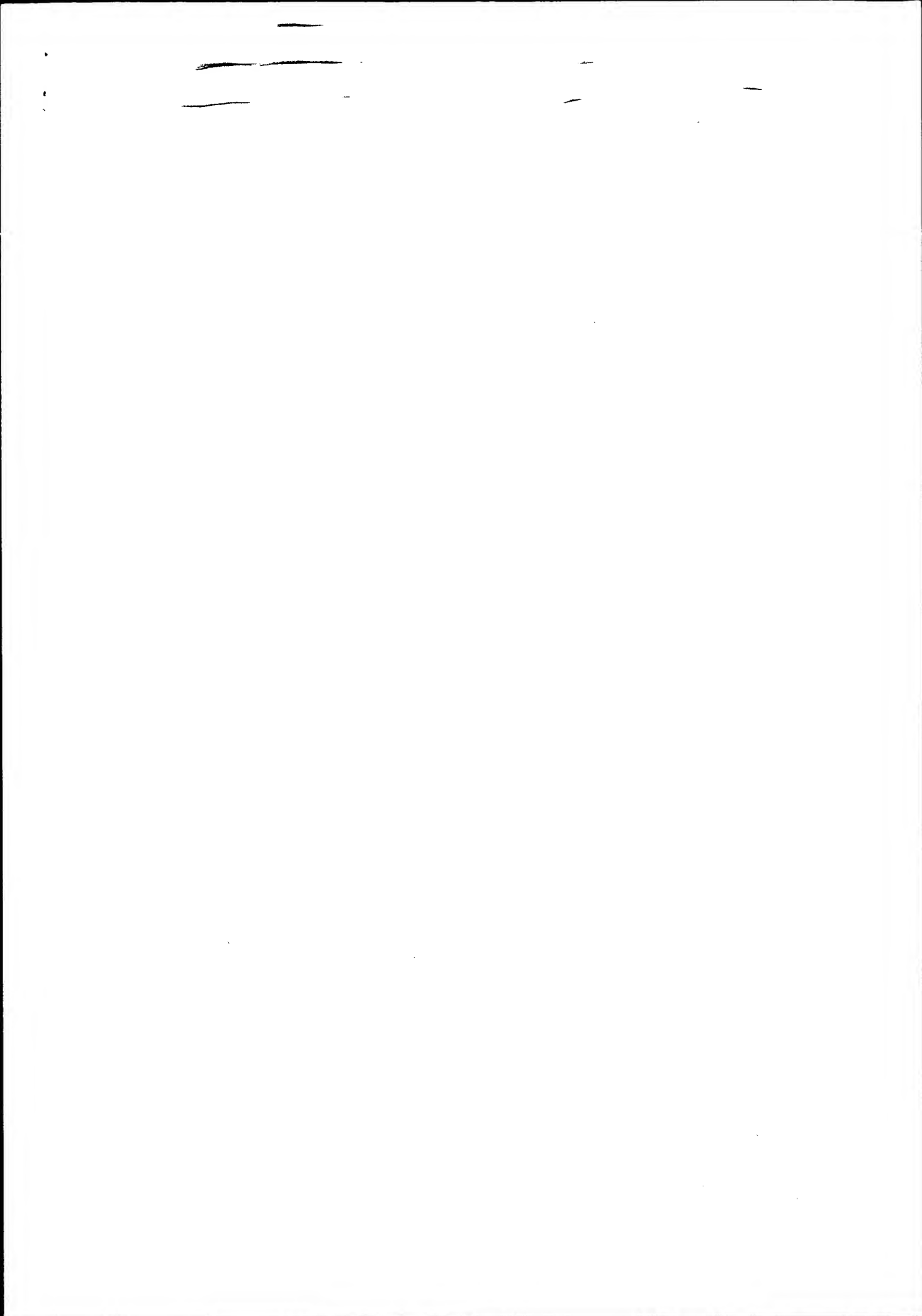




Fri Mar 1 09:12:03 2002

us-09-668-119-3.rpt







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: February 28, 2002, 19:37:53 ; Search time 16.91 Seconds  
(without alignments)  
1255.407 Million cell updates/sec

Title: US-09-668-119-3  
Perfect score: 3010  
Sequence: 1 MRKAGVAHKSXSKDMESHVF.....WPTSAHLSSTIPCTAHSPFP 579

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues  
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	383	12.7	708	1 GBF_DICDI	P36417 dictyostell
2	383	12.7	966	1 SSNG_YEAST	P14922 saccharomyc
3	378	12.6	1596	1 MAM_DROME	P21519 drosophila
4	366	12.2	1589	1 PHP_DROME	P39769 drosophila
5	327.5	10.9	1008	1 GALY_KLUTA	P32257 kluyveromyc
6	327.5	10.9	2038	1 FSH_DROME	P13709 drosophila
7	326.5	10.8	905	1 SNF5_YEAST	P18480 saccharomyc
8	310	10.3	255	1 LP61_EIMTE	P15714 elmeria ten
9	310	10.3	2124	1 Y192_HUMAN	O93074 homo sapien
10	308	10.2	356	1 GLT4_WHEAT	P10385 triticum ae
11	298.5	9.9	838	1 GLT4_WHEAT	P08489 triticum ae
12	298	9.9	1403	1 PRO_DROME	P29617 drosophila
13	296.5	9.9	648	1 KAPC_DICDI	P34099 dictyostell
14	296	9.8	839	1 GLT5_WHEAT	P10388 triticum ae
15	295.5	9.8	1123	1 DC11_DROME	P18169 drosophila
16	294	9.8	519	1 ELAV_DROVI	P23241 drosophila
17	294	9.8	2442	1 CBP_HUMAN	O92793 homo sapien
18	292	9.7	1589	1 DC13_DROME	P18171 drosophila
19	290.5	9.7	738	1 YRF4_YEAST	P35732 saccharomyc
20	285	9.5	1231	1 INVO_MOUSE	P48997 mus musculu
21	284.5	9.5	1231	1 YKT3_CAEEL	P34314 caenorhabdi
22	284.5	9.5	1902	1 SMF1_HUMAN	O14497 homo sapien
23	283	9.4	313	1 GDA7_WHEAT	P04727 triticum ae
24	283	9.4	319	1 GDA5_WHEAT	P04725 triticum ae
25	283	9.4	2441	1 CBP_MOUSE	P45481 mus musculu
26	282	9.4	848	1 YBVB_YEAST	P38266 saccharomyc
27	282	9.4	1638	1 BRM_DROME	P25439 drosophila
28	279.5	9.3	493	1 HMA8_DROME	P09087 drosophila
29	279.5	9.3	550	1 CCE_DROME	P41004 drosophila
30	278.5	9.3	660	1 GLT3_WHEAT	P08488 triticum ae
31	278.5	9.3	950	1 DC12_DROME	P18170 drosophila
32	274	9.1	648	1 GLT0_WHEAT	P10387 triticum ae
33	273.5	9.1	3828	1 TRX_DROVI	O24742 drosophila

34	273	9.1	1028	1 OVO_DROME	P51521 drosophila
35	272.5	9.1	618	1 ZESP_DROVI	O24762 drosophila
36	271.5	9.0	347	1 INVO_PIG	P18175 sus scrofa
37	270.5	9.0	395	1 SRY_MOUSE	O05738 mus musculu
38	269.5	9.0	1331	1 NFY3_HUMAN	O94916 homo sapien
39	267.5	8.9	700	1 BIB_DROME	P23645 drosophila
40	264	8.8	468	1 YOC1_CAEEL	O09260 caenorhabdi
41	264	8.8	1023	1 CLOC_DROME	O61735 drosophila
42	264	8.8	1443	1 E75C_DROME	P13055 drosophila
43	261.5	8.7	1081	1 GALY_YEAST	P19659 saccharomyc
44	256.5	8.5	1556	1 PRO_DROVI	O09641 drosophila
45	256	8.5	2414	1 P300_HUMAN	O09472 homo sapien

## ALIGNMENTS

RESULT	ID	GBF_DICDI	STANDARD	PRT	708 AA.
AC	P36417	GBF_DICDI			
DT	01-JUN-1994	(Rel. 29, Created)			
DT	01-JUN-1994	(Rel. 29, Last sequence update)			
DT	01-NOV-1997	(Rel. 35, Last annotation update)			
DE	G-BOX BINDING FACTOR (GBF).				
CN	GBFA.				
OS	Dictyostelium discoideum (Slime mold).				
OC	Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.				
OX	NCBI_TaxID=44689;				
RN	[1]				
RC	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
RC	STRAIN=AX3;				
RX	MEDLINE=94170994; PubMed=8125261;				
RA	Schmitzler G.R., Fischer W.H., Firtel R.A.;				
RT	"Cloning and characterization of the G-box binding factor, an				
RT	essential component of the developmental switch between early and				
RT	late development in Dictyostelium.";				
RL	Genes Dev. 8:502-514(1994).				
CC	-!- FUNCTION: CAMP-RESPONSIVE TRANSCRIPTIONAL ACTIVATOR REGULATING				
CC	LATE GENE EXPRESSION. ESSENTIAL COMPONENT OF THE DEVELOPMENTAL				
CC	SWITCH BETWEEN EARLY AND LATE DEVELOPMENT. BINDS TO A NUMBER OF				
CC	CA/GT-RICH GENE REGULATORY ELEMENTS.				
CC	-!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).				
CC					
CC	this SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC					
DR	EMBL; L29075; AAA21021.1; -.				
DR	TRANSFAC; T00315; -.				
DR	Dictydb; DD02046; gbfa.				
KW	Transcription regulation; Activator; DNA-binding; Nuclear protein;				
KW	Repeat.				
FT	REPEAT	339	368	1.	
FT	REPEAT	481	510	2.	
FT	DOMAIN	11	21	POLY-SER.	
FT	DOMAIN	115	263	GLN-RICH.	
FT	DOMAIN	270	292	POLY-RICH.	
FT	DOMAIN	549	557	POLY-ASN.	
FT	SEQUENCE	708 AA;	79268 MW;	B4BDBF04FACACCA CRC64;	
QY	Query Match	12.7%;	Score 383;	DB 1;	Length 708;
QY	Best Local Similarity	24.4%;	Pred. No. 1.5e-09;		
QY	Matches 170;	Conservative 57;	Mismatches 200;	Indels 270;	Gaps 23;
QY	44	HNKSSQASVSPPMNALOSLTGGPAAGAGTGMPRG-----	PGQS-LGG	86	
QY	6	HHQGNSSSSSSSSPSPTIGGSLNSITSLPLPSPISFTTAQNMNPPILEPPTSLDGG	65		







```

OY 189 QQQQHLIKLHNOQQQIQQQQQLRIAGLQQQQQQQQQQQQALENPPIQPP 248
DB 567 QQQQQQ-QQQQQQQQQQQQQQQQQQLPRLPQQQIQQAGVSYQMLNPPQGGPYTIPVIAH 624
OY 249 MOOP-----QPPPSQALP-----QOLOQHHPHHPPOPPPOOP--PVAONOPSQ--LPPOS 296
DB 625 QIQPFSTQAMHPPQSSQLPPOQQQQLQSVQHPPQQLQGGQQAQAPPLIGHNPPQNVLPKR 684
OY 297 QTPPLV-SQAQALPQOMLYTOPPLKFNAP--MVYQPP-----VDPQVQQQQTAVOTA 347
DB 685 YMEGAIHTLVDAVSSSTHTENTKSPRQPTHAIPQAATGITNAEPQVKKKLNPSNS 744
OY 348 QAAQVAPVQVQSSQLPMLSPSPG-----QQVQTP--QSMPP--PPQSPQPG 393
DB 745 NINKLVNTATSIENAKSEVSNQSPAVVESNTNNTSOEKKPVKANSIPVIGAOEPPQEA 804
OY 394 QPSSQPSNVSSGPAAPSPFLP--SPSPQSPQVPTARTPON-----FSVP-- 438
DB 805 SPPEATKAASVSPSTKPLNTEPESSVQPTVSSSSTTKANDOSTAETIELSTAVPAE 864
OY 439 -SP-----GPLNTPVNPSSVWSPAGSSQAEQQ-----YLD 468
DB 865 ASPVEDEVQHSKEGTEASAPSTEEAPASRDAEKQDETAATTITVIKPTLEME 924
OY 469 KIKQLSKYTE-----PLRRINKDKNED 492
DB 925 TYKEAKMREEDQTSQEKSPQENTLPRENVVQVEEDEN 963

RESULT 3
MAM_DROME STANDARD: PRT; 1596 AA.
AC 1 P21519;
DB 01-MAY-1991 (Rel. 18, Created)
DB 01-MAY-1991 (Rel. 18, Last sequence update)
DB 01-MAR-1992 (Rel. 21, Last annotation update)
DE NEUROGENIC PROTEIN MASTERMIND.
GN MAM.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN 111
RP SEQUENCE FROM N.A.
RC STRAIN-CANTON-S;
RX MEDLINE=91065516; PubMed=1701150;
RA Smoller D., Friedel C., Schmid A., Bettler D., Lam L.,
RA Yedvobnick B.;
RT "The Drosophila neurogenic locus mastermind encodes a nuclear protein
RT unusually rich in amino acid homopolymers.";
RL Genes Dev. 4:1688-1700(1990).
CC -1- FUNCTION: MAY HAVE A REGULATORY FUNCTION POSSIBLY IN ASSOCIATION
CC WITH THE N GENE PRODUCT.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DEVELOPMENTAL STAGE: DURING EARLY NEUROGENESIS MAM PRODUCTS ARE
CC UBQUITOUSLY LOCATED. DURING LATER STAGES THEY ACCUMULATE IN THE
CC CENTRAL NERVOUS SYSTEM.
CC -1- MISCELLANEOUS: THE PROTEIN HAS MANY AA HOMOPOLYMERIC DOMAINS: 21
CC POLY-GLN RUNS (FROM 5 TO 16 AA IN LENGTH), 4 POLY-GLY (6 TO 10
CC AA), 3 POLY-ASN (3 X 5 AA), 1 POLY-ALA (10 AA) AND 1 POLY-THR (5
CC AA) RUNS.
CC -1- SIMILARITY: TO OTHER NUCLEAR PROTEINS OF DROSOPHILA, TO CERTAIN
CC YEAST AND MAMMALIAN REGULATORY PROTEINS.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
```

```

CC -----
DB EMBL: X54251; CAA38152.1; -
DB PIR: A33106; A33106.
DB PIR: A36391; A36391.
DB FlyBase: FBgn0002643; mam.
KW Neurogenesis; Nuclear protein; Repeat.
FT DOMAIN 20 84 ARG/LYS-RICH (BASIC).
FT DOMAIN 127 190 GLN-RICH.
FT DOMAIN 196 219 GLN-RICH.
FT DOMAIN 259 304 ASN-RICH.
FT DOMAIN 355 388 GLY/ASN-RICH.
FT DOMAIN 392 406 GLN-RICH.
FT DOMAIN 407 440 GLY-RICH.
FT DOMAIN 651 671 GLN-RICH.
FT DOMAIN 700 714 GLN-RICH.
FT DOMAIN 759 816 GLN-RICH.
FT DOMAIN 987 996 ALA-RICH.
FT DOMAIN 1060 1079 5 X 2 AA TANDEM REPEATS OF G-V.
FT DOMAIN 1092 1107 8 X 2 AA TANDEM REPEATS OF V-G.
FT DOMAIN 1237 1252 7 X 2 AA TANDEM REPEATS OF G-V.
FT DOMAIN 1492 1496 POLY-THR.
FT DOMAIN 1559 1592 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 1596 AA; 167717 MW; B944D86EF359D605 CRC64;

Query Match 12.68; Score 378; DB 1; Length 1596;
Best Local Similarity 30.28; Pred. No. 4; Ce-09;
Matches 157; Conservative 30; Mismatches 143; Indels 190; Gaps 24;

OY 46 KKSQASVSDPMNALQ-SLTGSPAGAGIGMPRPGQSLGMSFGAMGQPMSLSGQP- 103
DB 537 KVEPPNODLINSLVKSESGLGHGFGGLGIDNPGMKMG-----GNPGNGSGFPN 589
OY 104 -PPTSGMAP-----SMAVSTAPPTQQLQQLQVLAQ----- 135
DB 590 GPNGTGAGPAPNAGNGSGNLMSEHPLAQTLLKQMAEQHKNKMGSGFRRPRHGM 649
OY 136 -QQQQQQQFQQQQQQAALQQ-----QQQQQQQQQF 163
DB 650 PQQQQQQQQQQQQQAQQQHGQMGQGPGRYNDYGGFPRDFGLGPRGQQQQQAQQ 709
OY 164 QAQGSAM-----QQQFQAVVQ-----QQQQLQQQQQQ 190
DB 710 QPQQQHLPQFHQKGRPRGAGMNVQNFIDIKQLFLFSSQGNDFDLKRLQQQQQAQQQQ 769
OY 191 QQHLLIKLHNOQQQIQ-----QQQQQLQRIAGLQQLQQQQQQQQQQQQALQEAQ--- 241
DB 770 QQ-----HHQQQQPRMGVPRNFKNQQQQQVPPQQQLQQQQQQQQQQQQDYSPFSNQ 824
OY 242 -----PP-----IQPP-----PMQPPRPPSGLPQQLQDMHT-----QHQP----- 275
DB 825 PMAANPLNCPRRGPGNGNQPPGNIQAQQQQP-----GAQPPQQQQRGNAGQGNPNPTGP 881
OY 276 -----PPQPPPVAVQNPQSPQLPQSPQTPPL-VSQ-----AQLPQOMLYTOPPLKFN 323
DB 882 GGNTPNAPQQQQQQQQSTTTTL-QMKQTQLHISQGGAGCIQVSAQGHLLSGDMK-SN 939
OY 324 APNVVQPPV-QPPVQQQQTAVQTAAQAVAPVQVQSSQLPMLSPSPSGQGVQTPPOS 381
DB 940 VSAVAAQQQVFPQQQAQQQQ-----QQQQQPG-----GTGPNPQQQQQQQPPHG 981
OY 382 MPP-----PPQSPQPPQSPQSPNSVSGPAPS 410
DB 982 GNAGGQVGVGVGVGNGCPNPGQQQQQPPNQMNSNANVPS 1021

RESULT 4
MAM_DROME STANDARD: PRT; 1589 AA.
AC P39769;
DB 01-FEB-1995 (Rel. 31, Created)
DB 01-FEB-1995 (Rel. 31, Last sequence update)
DB 20-AUG-2001 (Rel. 40, Last annotation update)
```



POLYHOMEOTIC-PROXIMAL CHROMATIN PROTEIN.  
 PH-P.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Phlebotomus; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 RX MEDLINE=92146957; PubMed=1346609;  
 RA Pechamillis M., Cheny N.S., Pierre D., Brock H.W.;  
 RT The polyhomeotic gene of Drosophila encodes a chromatin protein that  
 shares polyene chromosome-binding sites with Polycomb.;  
 RL Genes Dev. 6:223-233(1992).  
 [2]  
 RX MEDLINE=92039031; PubMed=1937015;  
 RA Peatrick J., Daly M., Randscholt N.B., Brock H.W.;  
 RT The complex genetic locus polyhomeotic in Drosophila melanogaster  
 potentially encodes two homologous zinc-finger proteins.;  
 RL Gene 105:185-195(1991).  
 CC -1- FUNCTION: BINDS TO POLYENE CHROMOSOMES. SEEMS TO INTERACT WITH  
 PC. MAY INTERACT WITH PROTEINS ALREADY BOUND TO PROMOTER  
 COMPLEXES AND MAY BE A NEGATIVE REGULATOR OF HOMEOTIC AND  
 SEGMENTATION GENES. PLAYS A ROLE IN REGULATING THE EXPRESSION OF  
 OTHER PAIR-RULE GENES SUCH AS EVE, FTZ, AND H.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- TISSUE SPECIFICITY: SALIVARY GLANDS.  
 CC -1- SIMILARITY: COMPAINS 1 SAM DOMAIN.  
 CC -1- SIMILARITY: TO MOUSE EARLY DEVELOPMENT REGULATOR PROTEIN RAE-28.  
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-9 IS THE INITIATOR.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: X63672; CAA4531.1;  
 DR EMBL: M64750; -; NOT\_ANNOTATED\_CDS.  
 DR PIR: S23632; S23632.  
 DR FlyBase: FBgn0004861; ph-P.  
 DR InterPro: IPR001660; SAM.  
 DR Pfam: PF00556; SAM.1.  
 DR SMART: SM00454; SAM.1.  
 DR zinc-finger; SAM.1.  
 FT DOMAIN 74  
 FT FT 40  
 FT FT 411  
 FT FT 450  
 FT FT 494  
 FT FT 494  
 FT FT 619  
 FT FT 630  
 FT FT 775  
 FT FT 775  
 FT FT 1233  
 FT FT 1233  
 FT FT 1365  
 FT FT 1387  
 FT FT 254  
 FT FT 254  
 FT FT 1415  
 FT FT 1415  
 FT FT 1589 AA; 167297 MW; A6DF0CF9106E1891 CRC64;  
 SQ SEQUENCE  
 Query Match 12.2%; Score 366; DB 1; Length 1589;  
 Best local similarity 29.7%; Pred. No. 1.4e-08;  
 Matches 158; Conservative 57; Mismatches 177; Indels 140; Gaps 24;  
 QY 7 ASKSSKDMESHVETKAKTRDEX-----LSLYA---RLIHFRRIHNKKSQA 50  
 DB 674 ASSVSTGTQNOOSLIKAKKRNKOQVPPALALIKTEIGVAGQNKVGHLLTTVQQQQAAT 733  
 QY 51 SVSDPMNNAQSLTGCPAAGAGTGMPPKPGGSGISGSGSGAMGQPMSTSGQPPETSGM 110  
 DB 734 NIQQVYN-----AAGNKM-----VYMSSTGTPTLQNGO 762

111 APHMAVSTATPOTOLQ-----QVVALQOQOQOQF-----QOQOQALQOQOQO-----Q 159  
 DB 763 TLHATAGVDEKQO 822  
 QY 160 QQQPQA-QQSAQQOQFQAVVQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO 200  
 DB 823 QQQQVNAQQQQAQVAAQQAQQAQQAQQAQQAQQAQQAQQAQQAQQAQQAQQAQQAQQAQ 882  
 QY 201 --NQO 252  
 DB 883 TSHQO 939  
 QY 253 QPPPSQALPQOQOQHHTQHPQPPQOQPPV-----AQNPQSLPQSQOTQPLVQAQA 307  
 DB 940 ---SGATSQOQTSQOQ---QHQQ-SGQLQSSVPSFVSSTTPAGTATSSALQALASAGA 992  
 QY 308 L-----PGQMYTOPPLKFFV-----RAPMVQOQPPVQOQOQOQOQOQOQOQOQOQO 351  
 DB 993 IQGTAKPGTCSSSPTSSVTTTNGSSPTVTS--TVASIQQAQI--QSAQVHQHQLI 1048  
 QY 352 --MVAQVQVSSSLPMLSSPSRGOVQTPQSMPPPPQPPQPPQPPQPPQPPQPPQPPQPP 409  
 DB 1049 SATIAGTD-----QPPQPSLPTTTPILAMTSMANATVGHLSIAPV 1093  
 QY 410 SPSSFLPSPSPQSPSPV--TATPQNPSPSPGPLNTP-VNPSVMSPAS 458  
 DB 1094 TVSVSTAVTSSPGQLVLTASSGGGSIPTATPKTPSKGPTATVPIQS 1145  
 RESULT 5  
 ID GALLY\_KLUJA STANDARD; PRT: 1008 AA.  
 AC P32257;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE TRANSCRIPTION REGULATOR PROTEIN GALL1.  
 GN GALL1.  
 OS Kluyveromyces fragilis (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
 NCBI\_TaxID=28985;  
 FT SEQUENCE FROM N.A.  
 FT MEDLINE=92020226; PubMed=1923818;  
 FT Dickson R.C., Hopper J., Mylin L.W., Gerardot C.J.;  
 FT "Sequence conservation in the Saccharomyces and Kluyveromyces GALL1  
 FT transcription activators suggests functional domains.";  
 FT Nucleic Acids Res. 19:5345-5350(1991).  
 CC -1- FUNCTION: AUXILIARY TRANSCRIPTION ACTIVATOR FOR GENES ENCODING  
 CC GALACTOSE-METABOLIZING ENZYMES. ESSENTIAL FOR NORMAL GROWTH ON  
 CC MISCELLANEOUS CARBON SOURCES, FOR SPOULATION AND MATING.  
 CC WITH GALL THAT HAS THE CAPACITY TO BIND DNA. ASSOCIATION BETWEEN  
 CC GALL1 AND GALL4 MAY SERVE TO EXPEDITE PHOSPHORYLATION OF GALL4.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: M68870; AAA35254.1;  
 DR Transcription regulation; Activator; Galactose metabolism.  
 FT DOMAIN 75  
 FT FT 84  
 FT FT 257  
 FT FT 287  
 FT FT 364  
 FT FT 411  
 FT FT 456  
 FT FT 464  
 FT FT 1008 AA; 114831 MW; 436DIEBAEAI77DB CRC64;  
 SQ SEQUENCE



```

RL Dev. Biol. 134:246-257(1989).
CC -!- FUNCTION: REQUIRED MATERIALIZED FOR PROPER EXPRESSION OF OTHER
CC HOMEOIC GENES INVOLVED IN PATTERN FORMATION, SUCH AS UBX.
CC -!- SIMILARITY: HIGH, TO HUMAN RING3 PROTEIN.
CC -!- SIMILARITY: CONTAINS 2 BROMODOMAINS.
CC -!- SIMILARITY: CONTAINS 1 ET DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M23221; AAA28540.1; -.
DR EMBL; M23222; AAA28541.1; ALT_TERM.
DR EMBL; M15762; AAA70424.1; -.
DR EMBL; M15763; AAA70423.1; -.
DR EMBL; M15764; AAA70422.1; -.
DR PIR; A43742; A43742.
DR HSSP; P04002; IMFA.
DR FLYBASE; FBgn0004656; fs(1)h.
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; Bromodomain; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SMO0297; BROMO; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; 2.
DR PROSITE; PS50014; BROMODOMAIN_2; 2.
KW Developmental protein; Bromodomain; Transmembrane; Repeat.
FT DOMAIN 51 123 BROMODOMAIN 1.
FT DOMAIN 495 567 BROMODOMAIN 2.
FT DOMAIN 945 1106 ET DOMAIN.
FT TRANSMEM 330 350 POTENTIAL.
FT TRANSMEM 451 471 POTENTIAL.
FT TRANSMEM 750 770 POTENTIAL.
FT TRANSMEM 790 810 POTENTIAL.
FT TRANSMEM 816 830 POTENTIAL.
FT TRANSMEM 874 894 POTENTIAL.
FT TRANSMEM 1731 1751 POTENTIAL.
FT TRANSMEM 1939 1959 POTENTIAL.
FT VARIANT 909 909 G->A.
FT VARIANT 1022 1022 H->RNPYY.
SQ SEQUENCE 2038 AA; 205332 MW; 849E0706D50A0098 CRC64;

Query Match 10.9%; Score 327.5; DB 1; Length 2038;
Best Local Similarity 26.8%; Pred. No. 6.3e-07;
Matches 150; Conservative 54; Mismatches 192; Indels 163; Gaps 20;

QY 44 HNKK-----SOASVSDPPNNAL--OSLTGCPAGA-----AGTGMPRR 78
   ||| | :||| | | | | |
Db 1226 HNKGPNRLSKVGGRPINALLPRISFGAGTAIYAATSOGSGIRLASMLHNRSGIG---- 1281

QY 79 GRGGSGLG-----GMSFGAMQRPMSLSG-----OPRGTSGMASHMAVYST 120
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1282 --GGDLGEHNAAIAALNLSGINSTGTACGGINNNGSGNNANPLGGSHGDAMVNVAASLSL 1339

QY 121 ATPOTQLDLGVVALIQDQQDQDF-----QQ--QQQAALDQQDQDQDQDQ 162
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1340 ASGKQITRFDDRDVEOVLASFSEFSGATGKSGLIDNFIMQCHLMRAPRPODQDQDQDQ 1399

QY 163 FOAQDSAMQDGFQAVLVUQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQ 219
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1400 FGHDQD-----QDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQ 1449

QY 220 LQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQ 255
   . :| :| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1450 MAAYLVQRNRRQDQDQAHNNGFNVADEFGMAERFDGIINTMASFRLDERSLDQDQDQ 1505

QY 256 PSQALPDQLDMNTGNHNPVRQDQDQDQDQDQDQDQDQDQDQDQDQDQDQ 315
   . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Db 1506 -----MLODQHNRQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQ 1546

```















OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_taxid=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92069760; PubMed=1720353;  
 RA Vaessin H., Grell E., Wolff E., Bier E., Jan L.Y., Jan Y.N.;  
 RT "Prospero is expressed in neuronal precursors and encodes a nuclear  
 RT protein that is involved in the control of axonal outgrowth in  
 RT Drosophila."  
 RL Cell 67:941-953(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92171948; PubMed=1540176;  
 RA Matsuzaki F., Koizumi K., Hama C., Yoshioka T., Nabeshima Y.;  
 RT "Cloning of the Drosophila prospero gene and its expression in  
 RT ganglion mother cells."  
 RL Biochem. Biophys. Res. Commun. 182:1326-1332(1992).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93083413; PubMed=1842358;  
 RA Chu-Iragui Q., Wright D.M., McNeill L.K., Doe C.Q.;  
 RT "The prospero gene encodes a divergent homeodomain protein that  
 RT controls neuronal identity in Drosophila."  
 RL Development Suppl. 2:79-85(1991).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM PROS-L).  
 RX MEDLINE=20503846; PubMed=11051550;  
 RA Xu C., Kaufmann R.C., Zhang J., Klady S., Carthew R.W.;  
 RT "Overlapping activators and repressors delimit transcriptional  
 RT response to receptor tyrosine kinase signals in the Drosophila eye."  
 RL Cell 103:87-97(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORM PROS-L).  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Gelinkner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blake J.R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,  
 RA Balcer R.M., Basu A., Baxendale J., Bayraktiroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Foster G., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merklow G., Milshina N.V., Moadary C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Slater E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Teector C., Turner R., Venier E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).  
 RN [6]  
 RP SIMILARITY TO C. ELEGANS CEH-26.  
 RX MEDLINE=94212446; PubMed=7909177;  
 RA Buerklin T.R.;  
 RT "A Caenorhabditis elegans prospero homologue defines a novel domain."  
 RL Trends Biochem. Sci. 19:70-71(1994).  
 CC -1- FUNCTION: INVOLVED IN THE CONTROL OF OTHER NEURONAL PRECURSOR  
 CC GENES AS WELL AS AXONAL OUTGROWTH AND PATHFINDING OF NUMEROUS  
 CC CENTRAL AND PERIPHERAL NEURONS. IT IS PROBABLY GENERALLY REQUIRED  
 CC FOR PROPER NEURONAL DIFFERENTIATION OF MOST OR ALL NEURONS & THEIR  
 CC PRECURSORS IN THE CENTRAL AND PERIPHERAL NERVOUS SYSTEMS, BUT NOT  
 CC FOR THE SPECIFICATION OF IDENTITY. PROSPERO PROTEIN MAY REGULATE  
 CC TRANSCRIPTION BY BINDING TO DNA.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: PROS-L (SHOWN HERE) AND PROS-  
 CC S. ARE PRODUCED BY ALTERNATIVE SPLICING. THEY HAVE DIFFERENT N-  
 CC TERMINAL AMINO ACIDS OF THE HOMEODOMAIN.  
 CC -1- TISSUE SPECIFICITY: NEURONAL PRECURSORS. EXPRESSED IN THE  
 CC DEVELOPING CNS. LENS-SECRETING CONE CELLS OF THE EYE, AND MIDGUT.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN NEURONAL PRECURSORS EARLY DURING  
 CC FORMATION.  
 CC -1- SIMILARITY: BELONGS TO THE PROSPERO FAMILY OF HOMEBOX PROTEINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; M81389; AAA28841.1; -;  
 DR EMBL; D10609; BAA01464.1; -;  
 DR EMBL; Z11743; CAA77802.1; -;  
 DR EMBL; AF190403; AAF05703.1; -;  
 DR EMBL; AE003691; AAF54628.1; ALT. INTR.  
 DR PIR; A41089; A41089.  
 DR PIR; J01397; J01397.  
 DR FlyBase; FBgn0004595; pros.  
 KW Nuclear protein; Transcription regulation; DNA-binding; Homeobox;  
 KW Developmental protein; Alternative splicing.  
 FT DOMAIN 4 12  
 FT DOMAIN 28 31  
 FT DOMAIN 32 35  
 FT DOMAIN 188 191  
 FT DOMAIN 216 264  
 FT DOMAIN 270 286  
 FT DOMAIN 318 354  
 FT DOMAIN 431 437  
 FT DOMAIN 505 508  
 FT DOMAIN 700 1048  
 FT DOMAIN 934 937  
 FT DOMAIN 991 998  
 FT DOMAIN 1074 1082  
 FT DOMAIN 1127 1137  
 FT DNA\_BIND 1241 1303  
 FT DNA\_BIND 1304 1403  
 FT VARSPLIT 1216 1244  
 FT CONFLICT 76 98  
 FT CONFLICT 120 144  
 FT CONFLICT 418 418  
 FT CONFLICT 677 677  
 FT CONFLICT 802 802  
 FT CONFLICT 958 958  
 FT CONFLICT 1048 1048  
 SQ SEQUENCE 1403 AA; 153569 MW; 9EF9973E24E238E CRC64;

Query Match 9.9%; Score 298; DB 1; Length 1403;  
 Best Local Similarity 22.9%; Pred. No. 7.4e-06;











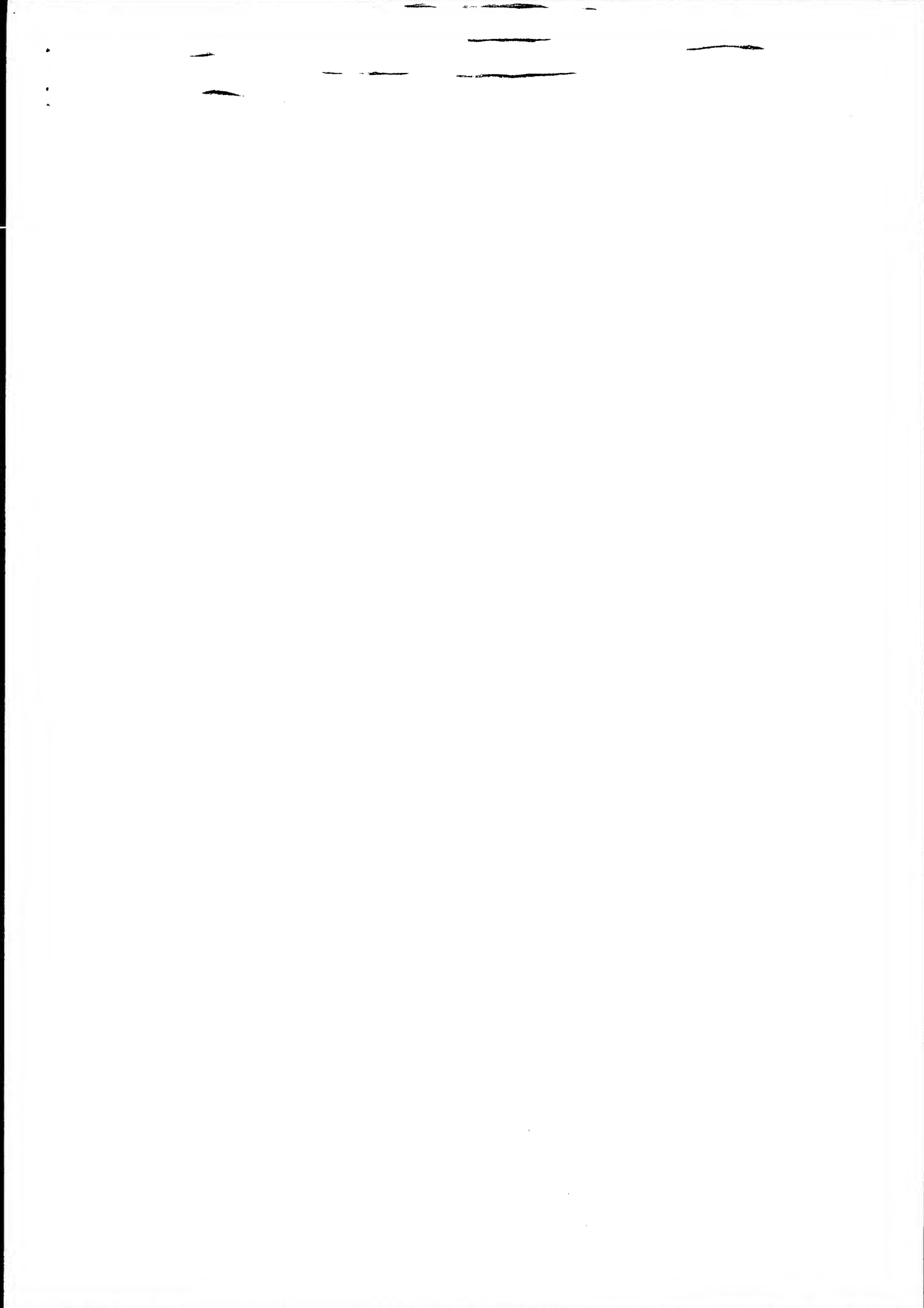




Fri Mar 1 09:12:03 2002

us-09-668-119-3.rsp







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

protein - protein search, using sw model

Run on: February 28, 2002, 19:36:28 ; Search time 26.64 Seconds

(without alignments)  
3179.117 Million cell updates/sec

Title: US-09-668-119-3

Perfect score: 3010

Sequence: 1 MKRAGVAHSKSSKDMESHV.....WPTSAHLSRIPTAHSPF 579

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 10%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_17:\*  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp rodent:\*  
12: sp.virus:\*  
13: sp.vertebrate:\*  
14: sp.unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3010	100.0	579	4 Q9P1T3	Q9P1T3 homo sapien
2	1780	59.1	349	4 O15413	O15413 homo sapien
3	469.5	15.6	749	5 Q9Y149	Q9Y149 drosophila
4	460.5	15.3	1069	5 Q9VF64	Q9VF64 drosophila
5	456.5	15.2	2123	5 Q909S7	Q909S7 dictyostell
6	447	14.9	796	5 Q45322	Q45322 caenorhabdi
7	439	14.6	4957	4 O14687	O14687 homo sapien
8	439	14.6	5262	4 O14686	O14686 homo sapien
9	423	14.1	2285	5 Q9VP17	Q9VP17 drosophila
10	419	13.9	1542	5 Q9VE07	Q9VE07 drosophila
11	415.5	13.8	1798	5 Q9VUB6	Q9VUB6 drosophila
12	414.5	13.8	900	5 Q9VK09	Q9VK09 drosophila
13	400	13.3	1893	5 Q24279	Q24279 drosophila
14	399	13.3	1893	5 Q24279	Q24279 drosophila
15	398	13.2	926	5 Q9W3G1	Q9W3G1 drosophila
16	397	13.2	2251	5 Q22190	Q22190 caenorhabdi
17	395.5	13.1	1013	5 Q9VY77	Q9VY77 drosophila
18	395.5	13.1	2703	5 Q9VPL9	Q9VPL9 drosophila
19	395.5	13.1	5322	5 Q9N164	Q9N164 drosophila

20	394.5	13.1	1954	5 Q9VN82	Q9VN82 drosophila
21	389.5	12.9	5533	5 Q9UC63	Q9UC63 drosophila
22	383.5	12.7	1012	11 Q64028	Q64028 mus musculu
23	383.5	12.7	1366	5 Q9V6W8	Q9V6W8 drosophila
24	383.5	12.7	1594	5 Q9V6W7	Q9V6W7 drosophila
25	382.5	12.7	5476	5 Q9N017	Q9N017 drosophila
26	382.5	12.7	5533	5 Q9VPL2	Q9VPL2 drosophila
27	382.5	12.7	5554	5 Q9NHN1	Q9NHN1 drosophila
28	382.5	12.7	5560	5 Q9VPL1	Q9VPL1 drosophila
29	380.5	12.6	1004	4 P78364	P78364 homo sapien
30	379	12.6	5539	6 Q9GM67	Q9GM67 equus caball
31	377	12.5	1655	5 Q24754	Q24754 drosophila
32	374	12.4	1164	10 Q9SY06	Q9SY06 arbidopsis
33	372	12.4	1165	10 Q9SBK4	Q9SBK4 arbidopsis
34	372	12.4	1165	10 Q9LC25	Q9LC25 arbidopsis
35	372	12.4	1234	5 Q01505	Q01505 caenorhabdi
36	367.5	12.2	929	5 Q9BLX2	Q9BLX2 dictyostell
37	367	12.2	1359	5 Q9NF31	Q9NF31 drosophila
38	367	12.2	1624	5 Q9W523	Q9W523 drosophila
39	365	12.1	1589	5 Q46097	Q46097 drosophila
40	365	12.1	4823	13 Q93321	Q93321 fuqu rubrip
41	362	12.0	3502	5 Q9VYJ9	Q9VYJ9 drosophila
42	360	12.0	2309	5 Q9W2U7	Q9W2U7 drosophila
43	358.5	11.9	3080	5 Q9VXR3	Q9VXR3 drosophila
44	357	11.9	2977	5 Q9VAP9	Q9VAP9 drosophila
45	354.5	11.8	1217	5 P91094	P91094 caenorhabdi

## ALIGNMENTS

RESULT 1  
Q9P1T3 PRELIMINARY; PRT; 579 AA.  
ID Q9P1T3  
AC Q9P1T3  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE TPA INDUCIBLE PROTEIN.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP  
RA Abraham S., Solomon W.B.;  
RT "A novel glutamine rich putative transcriptional adaptor protein  
RT preferentially expressed in placenta and bone marrow tissues."  
RT Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF056191; AAC12944.1; -  
DR InterPro: IPR002965; P. rich. extensn.  
DR PRINTS: PR01217; PRICHEXTENS.  
SQ SEQUENCE 579 AA; 63879 MW; CCC9E710C6BDD02F CRC64;

Query Match 100.0%; Score 3010; DB 4; Length 579;  
Best local similarity 100.0%; Pred. No. 1.9e-215;  
Matches 579; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRAGVAHSKSSKDMESHVFLKAKTRDEYLSIVARLIHFRIHKKKSQASVSPMNAIQ 60  
1 MKRAGVAHSKSSKDMESHVFLKAKTRDEYLSIVARLIHFRIHKKKSQASVSPMNAIQ 60  
QY 61 SITGSPAGAGAGIGMPGPGSGISGMSFGMGQPMISGQPPGTSGMAPHSAVYST 120  
61 SITGSPAGAGAGIGMPGPGSGISGMSFGMGQPMISGQPPGTSGMAPHSAVYST 120  
DB 61 SITGSPAGAGAGIGMPGPGSGISGMSFGMGQPMISGQPPGTSGMAPHSAVYST 120  
QY 121 ATPPTQLDQLQYVALDQ 180  
121 ATPPTQLDQLQYVALDQ 180  
DB 121 ATPPTQLDQLQYVALDQ 180  
QY 181 GQQLDQ 240  
GQQLDQ 240



[illegible]

RESULT	2
ID	015413
AC	015413
DT	01-JAN-1998 (TREMBLrel_05, Created)
DT	01-JAN-1998 (TREMBLrel_05, Last sequence update)
DT	01-JUN-2000 (TREMBLrel_14, Last annotation update)
DE	CRC67A (FRAGMENT).
GN	CRC67A.
OS	Homo Sapiens (Human)
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NBPI_TaxID=9606;
RX	[1]
RP	SEQUENCE FROM N.A.
RC	ISSUE GENERAL CORTEX;
RC	MEDLINE=97359492; PubMed=9225980;
RA	Margolis R.L., Abraham M.R., Gatchell S.B., Li S.H., Kidwai A.S.,
RA	Breschel T.S., Stine O.C., Callahan C., McNinis M.G., Ross C.A.,
RL	cDNAs With long CAG trinucleotide repeats from human brain.";
DR	Hum. Genet. 100:114-122(1997).
FT	EMBL: U080745; AAB91443.1; -.
FT	NON_TER I 1
SHDUTNCE	349 AA; 349 MW; D073F67E22B70533 CRC64;

[illegible]

	QY	Db	QY	Db
401	SNVSSGPPAPSSSFLPSSPPSSOPVYAPRPNQNSVSPSPPLPTTPVNPSSVMSPASSQ	4601	SNVSSGPPAPSSSFLPSSPPSSOPVYAPRPNQNSVSPSPPLPTTPVNPSSVMSPASSQ	4601
243	SNVSSGPPAPSSSFLPSSPPSSOPVYAPRPNQNSVSPSPPLPTTPVNPSSVMSPASSQ	3023	SNVSSGPPAPSSSFLPSSPPSSOPVYAPRPNQNSVSPSPPLPTTPVNPSSVMSPASSQ	3023
461	AEEQYITDKLKQLSYITPLPRMTINKIDKNDKDKKDJSKMKSLD	505	AEEQYITDKLKQLSYITPLPRMTINKIDKNDKDKKDJSKMKSLD	505
303	AEEQYITDKLKQLSYITPLPRMTINKIDKNDKDKKDJSKMKSLD	347	AEEQYITDKLKQLSYITPLPRMTINKIDKNDKDKKDJSKMKSLD	347

RESULT	3		
Q9Y149			
ID	Q9Y149	PRELIMINARY;	PRT; 749 AA.
AC	Q9Y149.		
DI	01-NOV-1999	(TrEMBLrel. 12, Created)	
DT	01-NOV-1999	(TrEMBLrel. 12, last sequence update)	
DT	01-JUN-2001	(TrEMBLrel. 17, last annotation update)	
DE	BCDNA:GHO3922 PROTEIN.		
GN	ACR105 OR BCDNA:GHO3922 OR CG4184.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
CC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
CC	Empidoidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
LN	[1]		
RP	SEQUENCE FROM N.A.		

NC  
RA STRAIN=BSR62KE1;  
RA MEDLINE=20136006; PubMed=107311132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanalides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.J., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.E., Richards L., Vandal M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Miller B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abdl J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Berens P.V., Bernier B.P., Bhandal D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadenot E., Center A., Chandra I.,  
RA Cherry J.M., Cusum M.S., Dahlke C., Davenport L.B., Davies P.,  
RA de Palacios B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup I.E., Downes M., Guzman R.A., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.C., Gelbart W.M., Glasser K.,  
RA Glodde A., Gong F., Gorelli J.H., Gu Z., Guan P., Harris N.L.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck T.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeigbam C.,  
RA Jasthi M., Kalush F., Kapran G.H., Ke Z., Kennison J.A., Kethum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.Y., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Munz D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Paclet J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Ramos I., Simpson M., Skupski M.P., Smith T.,  
RA Spher E.E., Spreading A.C., Stapleton M., Skupski M.P., Smith T.,  
RA Strickas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassaman D.A., Weinstock G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng X.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RA The genome sequence of *Drosophila melanogaster*.  
RL Science 287:2185-2195(2000).  
[2]

SEQUENCE FROM N.A.  
RA Rubin G.M., Wan K.H., Harvey D., Lewis S.E., Brokstein P., Tsang G.,  
RA Agbayani A., Accata T.T., Baxter E., Blazer R.G., Buehloff C.,  
RA Champe M., Chavez C., Chew M., Doyle C.M., Fattán D.E., Frise E.,  
RA Galle R., George R.A., Harris N.L., Hoskins R.A., Evans-Holm M.,  
RA Houston K.A., Hummasti S.R., Kim E., Li P., Moshrefi M., Paclet J.M.,











```

QY 372 PGGVOTPSGMPSPSPQGPSS--QPNNSVSGAPSPSSFLPSPSPQSPVTA 429
DB 609 PPMVATVTRGPPPT-----RPASVLAPESSV--PKSQPR---PSTTPPTTRATIR 657
QY 430 RPPONSVS-PGLPLTPVNPSSVMPASSQAEBO-----OYLDKLKOLSKYTEPLR 481
DB 658 PTPONKAVTSRAPAT-----SSVTPSPSQTQIRKPLAVLIDSNESTKTEILL 712
QY 482 R-----MINKIDKNEKDKLSKM 500
DB 713 RDTYGLPLTFYDKTKDPAIORLOQL 741

RESULT 7
ID 014687 PRELIMINARY; PRT; 4957 AA.
AC 014687;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ALR.
GN ALR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97388474; PubMed=9247308;
RA Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,
RA Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,
RA Canaan E.;
RT "Structure and expression pattern of human ALR, a novel gene with
RT strong homology to ALL-1 involved in acute leukemia and to Drosophila
RT trithorax."
RL Oncogene 15:549-560(1997).
DR EMBL: AF010404; AAC51735.1; -.
DR InterPro: IPR000910; HMG_12_box.
DR InterPro: IPR001214; SET.
DR InterPro: IPR001822; Recombinase.
DR InterPro: IPR001965; PHD.
DR InterPro: IPR002965; P-rich_extensn.
DR InterPro: IPR003616; PostSET.
DR InterPro: IPR003888; FYrich_N.
DR InterPro: IPR003889; FYrich_C.
DR Pfam: PF00628; PHD; 3.
DR Pfam: PF00856; SET; 1.
DR PRINTS: PR01217; PRICHEXTENS.
DR PROSITE: PS00398; RECOMBINASE_2; UNKNOWN_1.
DR PROSITE: PS00280; SET; 2.
DR SMART: SM00542; FYRC; 1.
DR SMART: SM00541; FYRN; 1.
DR SMART: SM00398; HMG; 1.
DR SMART: SM00249; PHD; 4.
DR SMART: SM00508; PostSET; 1.
DR SMART: SM00317; SET; 1.
SQ SEQUENCE 4957 AA; 531848 MW; 1026562E1419CEBD CRC64;

```

```

DB 3327 QQQQQQLQQQQQ---QQQLQQQQQLQQQQQLQQQQQLQQQQQLQQQQQLQQQQQL 3379
QY 217 QQLQLQQQQQQQQQQQ-----QQQQQALNAPPYQQPQQ----- 250
DB 3380 QQQQQQQQQQQQQQQQQQQQQQLNQSRTLLSPQQQQQQQVALGGMGPAKPLQHFSSPGALPTL 3439
QY 251 ----- 250
DB 3440 LITGEQNTVPNAVSSATEGPTTHGGPLAIGTTPESMATEPEYKPSLSGDSQLLVQ 3499
QY 251 -QPQPPSQ-----ALPQQLQ-----MH----- 268
DB 3500 PQQPQSSQLQPLRLRPGQQQQQVSLHTAGGSHGLGSSSSSEASSVPHLLAQSV 3559
QY 269 -----HTQHNDPPQPPQPPVAAQNSQLPQSQSTQPLVSAQALPQMLYTQPLK 320
DB 3560 SLGDQPGSMTQNLGPPQPPMLERPMQNTGPPQ--KGPVLAGGGILPG--VGIMPTVG 3615
QY 321 FYRAPH--VVGQPP-----VQPVQVQ-----QTAQVTAQAQAV-----APGVQS--QS 362
DB 3616 QLRALQGVLANPQLRHLSPQQQQQLQALLMQRLQDSQAVRQTPRYQEPGTQSPLOG 3675
QY 363 SL---PMLS--PSP-----GQVOTPSMPPPP-----QSPQPGQPS 396
DB 3676 LLGCPQLGPPGPPQGTPLQELGAGRPQGPRLPAPGALSTGVLGVPVHTP-----PP 3731
QY 397 SPPNSVSSGAPSPSSFLPSPSPQSPSPVTAQTPQNFVSP-----QSPQPGQPS 440
DB 3732 SSPQPKRPSQLPSPSSPLPTEALPPTNGTPK--PQGPTEPPRGVSPAAQADLTLE 3790
QY 441 ---GRLTPVNPSSVMPASS-----QAEEQYDLTKOLSKYTEP 479
DB 3791 SKGLGPPMPDNLATQKPESSSLVPGHLDVNGVYFEASQLSIKQEP 3839

RESULT 8
ID 014686 PRELIMINARY; PRT; 5262 AA.
AC 014686;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ALR.
GN ALR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97388474; PubMed=9247308;
RA Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,
RA Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,
RA Canaan E.;
RT "Structure and expression pattern of human ALR, a novel gene with
RT strong homology to ALL-1 involved in acute leukemia and to Drosophila
RT trithorax."
RL Oncogene 15:549-560(1997).
DR EMBL: AF010403; AAC51734.1; -.
DR InterPro: IPR000910; HMG_12_box.
DR InterPro: IPR001214; SET.
DR InterPro: IPR001822; Recombinase.
DR InterPro: IPR001841; Znf_ring.
DR InterPro: IPR001965; PHD.
DR InterPro: IPR002965; P-rich_extensn.
DR InterPro: IPR003616; PostSET.
DR InterPro: IPR003888; FYrich_N.
DR InterPro: IPR003889; FYrich_C.
DR Pfam: PF00628; PHD; 5.
DR Pfam: PF00856; SET; 1.
DR PRINTS: PR01217; PRICHEXTENS.
DR PROSITE: PS00398; RECOMBINASE_2; UNKNOWN_1.

```















RT "The genome sequence of *Drosophila melanogaster*."  
 (RL Science 287:2185-2195(2000)).  
 DR EMBL: AE003639; AAF53274.1;  
 DR FlyBase: FBgn0032497; CG6043.  
 DR InterPro: IPR002965; P-rich\_extensn.  
 DR PRINTS: PR01217; PRICHEXTENS.  
 SQ SEQUENCE 900 AA; 98948 MW; DD99C0057CDA5E44 CRC64;

Query Match 13.8%; Score 414.5; DB 5; Length 900;  
 Best Local Similarity 30.8%; Pred. No. 8.2e-23;  
 Matches 165; Conservative 42; Mismatches 192; Indels 137; Gaps 24;

```

QY 10 KSKDMESHVFLKAKTRDEYLSVRLILHFRDIHNKKSQASVSDPMNALDSTL-----G 64
DB 185 KVAPEVEKNTFMK-----DEYI-----KD-RKSKPLPEVAPRANGGSSSTSDAVD 229
QY 65 GPAAAGAGI--GMPRRGSGSLGMSFGAGMGPMUSLGGPRPGTSGMASHMAVSTAT 122
DB 230 GPRSSASVSSSYSPPTPQVPV-----AKSPVQYQPPRPA-----PQQQSQSQSQSQ 282
QY 123 PQTQLDQVAL-----PQQQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQ 149
DB 283 PATRPFERSVPMPTSPAVNYTRKSDSPSRPFEPQSQSQSQSQSQSQSQSQSQSQSQSQSQSQ 342
QY 150 ALQD-----PQQQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQ 199
DB 343 SPQQRPRPAISPLAQVQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQSQ 392
QY 200 QNQDQIQDQDQIQDQDQIQDQDQIQDQDQIQDQDQIQDQDQIQDQDQIQDQDQIQDQDQIQD 250
DB 393 -QDQDQFQDQDQIQDQDQIQDQDQIQDQDQIQDQDQIQDQDQIQDQDQIQDQDQIQDQDQIQD 448
QY 251 QPQRRPSALRQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQ 299
DB 449 QDQDQSPQPIYNNVQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQ 505
QY 300 -----PLVSAQALRPGMLYTPRLKFRAPMVVQPPVQVQVQVQVQVQVQVQVQVQVQVQVQ 354
DB 506 SLVIAFLAQPTERPAQRIILQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQ 565
QY 355 PGVQVQSS--LP-MLSSPSPGQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 410
DB 566 PMARLENGVNLPTLRTQTPQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQVQV 624
QY 411 PSSPLPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 465
DB 625 PVSAPPI-----SLQNFSGNPDGGLRLQIN-----LNTNGSSNNTNQ 663

```

## RESULT 13

Q9FR41 PRELIMINARY; PRT; 455 AA.  
 ID Q9FR41  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE SECALIN PRECURSOR.  
 GN GSEC2-1.  
 OS Secale cereale (Rye).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Triticeae; Secale.  
 OX NCBI\_TaxID=4550;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, IMPERIAL;  
 RA Murray F.R., Skeritt J.H., Appels R.;  
 RT "A gene from the Sec2 (G11-R2) locus of a wheat 2RS.2B1 chromosomal  
 translocation line."  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF201084; AAG35598.1;  
 DR InterPro: IPR001954; G11a-glutenin.

DR InterPro: IPR002965; P-rich\_extensn.  
 DR PRINTS: PR00208; GLIADGLUTEN.  
 DR PRINTS: PR01217; PRICHEXTENS.  
 FT CHAIN 20 455 SECALIN  
 SQ SEQUENCE 455 AA; 51731 MW; C1A1176DD045D927 CRC64;

Query Match 13.3%; Score 400; DB 10; Length 455;  
 Best Local Similarity 33.2%; Pred. No. 5e-22;  
 Matches 135; Conservative 39; Mismatches 175; Indels 58; Gaps 15;

```

QY 115 MAVVSTAT--PQTQLDQVALDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQD 167
DB 7 LAILAMATATATANNMGNPSGVQCPQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQ 66
QY 168 SAMQDQ-----FQAVVQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQ 223
DB 67 SSPQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQ 115
QY 224 QDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQ 280
DB 116 PQDQVQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQ 175
QY 281 QPVAQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQD 340
DB 176 QPQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQD 224
QY 341 QTVQVQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQ 397
DB 225 Q---PYDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQ 281
QY 398 QPNSNVSSGAPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 454
DB 282 QSQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQD 334
QY 455 -----PAGSQAEQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQ 495
DB 335 SKLPQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQ 381

```

## RESULT 14

Q24279 PRELIMINARY; PRT; 1893 AA.  
 ID Q24279  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE CNO.  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidae; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CANTON-S;  
 RX MEDLINE=95212910; PubMed=7698650;  
 RA Miyamoto H., Nihonmatsu I., Kondo S., Ueda R., Togashi S., Hirata K.,  
 Ikegami Y., Yamamoto D.;  
 RT "cno encodes a novel protein containing a G1GF/DHR motif and  
 functions with Notch and scabrous in common developmental pathways in  
*Drosophila*."  
 RL Genes Dev. 9:612-625(1995).  
 DR EMBL: D49534; BAA08478.1;  
 DR HSSP: Q12923; 3PDZ.  
 DR FlyBase: FBgn0000340; cno.  
 DR InterPro: IPR002710; DIL.  
 DR InterPro: IPR000253; FHA\_domain.  
 DR InterPro: IPR001478; PDZ.  
 DR InterPro: IPR000159; RA.  
 DR Pfam: PF01843; DIL; 1.  
 DR Pfam: PF00498; FHA; 1.

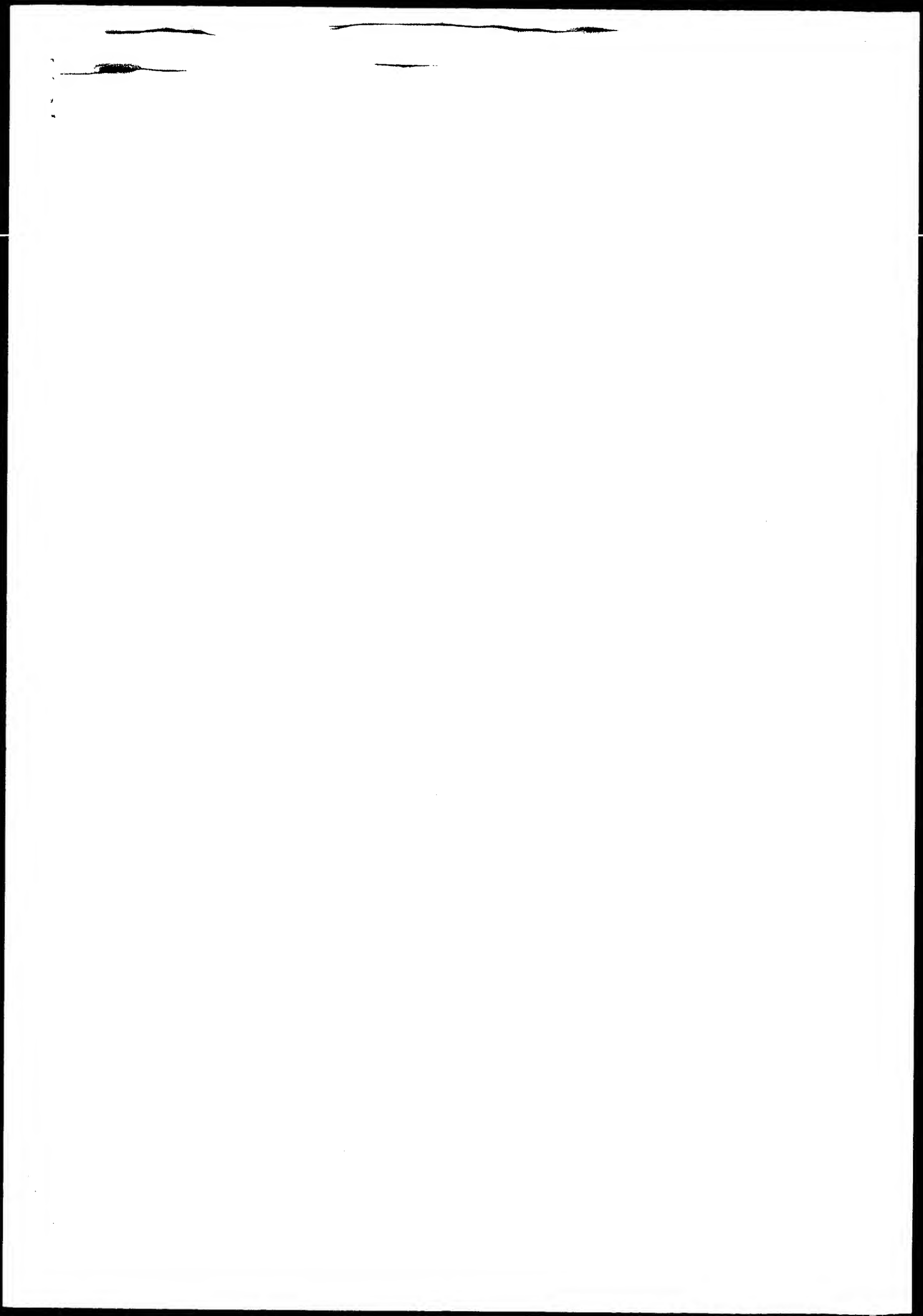






















/organism="Homo sapiens"

422 ProSerGlnSerProValThrAlaArgThrProGlnAsnPheSerValPr 438



900	CCTCCTCCAGACGCCAGTGCAGCGCCGCGAACC	CCCAAGAACTTCATGTGCCCC	94.9
438	OserProGlyProLeuAsnThrProValAnProSerSerValMetSerP		455
950	CTCACCTGGACCTTTAAACACACTGTGAACCCAGCTGTGTCAAGACC		999
455	roAlagIyserSerGlnAlaGlucInglngIntYrleuAspyLseuLys		471
1000	CAGCTGTGCTCAGCC/GCCTAGAGAGAAGCATGTGCAAGCTGAAG		1049
472	GlnleuSerItyTYrIlegluProleuArGArgNetilleasnyLileas		488
1050	CAGCTGTGGAAGTAC/TGAGACCCCCTCGCGCGCATGATCAACAAGATCGA		1099
488	pLyAsnGlnuAsparC LysILysAspieuSerIyMelLysSerleuleuA		505
1100	CAAGAACGAAACAG/AANAAGACCTGAGTAAGATGAAGAGCCCTCTGG		1149
505	splleleuThraspeOserLysArgCySproLeuLYrThleuGlnLys		521
1150	ACAATTGTACAGACCCCTCGAAGGGTGTCCTCGAAGACCTTCCAAG		1199
522	CysgGuilealaleuCluLysleuLysasnspMet.ArgcypProleup		538
1200	TGTGAGATCGCCCTGCAGAAACTCAAGATGACAMGGGGGTGCCACATCC		1249
538	rohLSARPROARGCYSHISARPROANSERSTHRTYALISERARG		554
1250	CCGACGCCGCCCGTCGCCACCGACCAACACAGACTACTATGCCAGCGCC		1299
555	SerTrpMetProSerTrpProThSerAlahisLeuSerTrpIlePr		571
1300	TCTGTGAATGGCTCC/GGCGCAACATCGCCTACCTGTTCACCATTC		1349
571	oCysThraAlahisSerPhedInPro		579
1350	CTGTACCGCACATTCGTTCCAGCCA		1374

seq\_name: gb\_est2:B6281933

seq_documentation_block:	LOCUS	DEFINITION	EST
B6281923	886 bp	mRNA	21-FEB-2001
602403104FI NIH_MGC_20 Homo sapiens cDNA clone IMAGE:454543			5'
mRNA sequence.			
B6281923			
B6281923.1 GI:13030849			
EST.			
human.			
Homo sapiens			
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Euarchontia; Primates; Carnivora; Homnidae; Homo.			
NIH-MGC http://mgc.ncl.nih.gov/			
National Institutes of Health, Mammalian Gene Collection (MGC)			
Unpublished. (1999)			
Contact: Robert Strausberg, Ph.D.			
Email: gsa@femail.nih.gov			
Tissue Procurement: ATCC/DCTP			
cDNA Library: Preparation: Ling Hong/Rubin Laboratory			
DNA Library: Arrayed by: The I.M.A.G.E. Consortium (LMNL)			
DNA Sequencing by: Ince Genomics, Inc			
Clone distribution: MGC clone distribution information can be			
found through the I.M.A.G.E. Consortium/LMNL at: image.liml.gov			
Plate: LMCL130 row: 9 column: 02			
High quality sequence stop: 876.			
Location/Qualifiers			
1. 886			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
/clone="IMAGE:454543"			
/clone_id="NIH_MGC_20"			
/tissue_type="melanotic melanoma"			

/lab host="DH10B (phage-resistant)"  
 /note="Organ: skin; Vector: pOTR1; Site\_1: XhoI; Site\_2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGCAAGAG(G). Size-selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."



```

602 TCCGACCCACGCTTGTGTCTCAGAGCGGACAGCTCTCCCTGGACAAAT 651
312 tleuTyThrGlnProLeuLysPheValArgAlaPro.MetValVal 328
||||| 328
652 GTGTATACCAACGACCAATGAAATTTGCGAGCTCTGATGTGTGTG 701
329 GlnGln.ProProVal.GlnProGlnValGlnGlnGlnGlnGlnGln 344
||||| 344
702 CAGCAGGCGCCAGTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTA 751
345 GlnThrAlaGlnAlaAlaGlnMetValAlaProGlyValGlnValSerG 361
||||| 361
752 CAGACCTGCTCAGGCTCCAGATGTGTGTGTGTGTGTGTGTGTGTGTGT 801
361 nSerSerLeuProMetLeu.....SerSerProSerProGlyGlnG 375
||||| 375
802 GGAAGCTTGGCCAGAGTGGGATGCACATAGACCCCGTCCGCTGACAA 851
375 lValGlnThrProGlnSerMetProProProProGlnProSerPro 390
||||| 390
852 CGCTG.....TGTCCGATCCCTCAAGCTCAATCCT 883

seq_name: gb_est1:AL046886
seq_documentation_block:
LOCUS AL046886 757 bp mRNA EST 29-FEB-2000
DEFINITION DKFZP586E2117.1 586 (synonym: hute1) Homo sapiens cDNA clone
AL046886
ACCESSION AL046886
VERSION AL046886.1 GI:5434945
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Koehler K., Beyer A., Mewes H.W., Gassenhuber J. and Wiemann S.
TITLE EST (Koehler, et al.)
JOURNAL Unpublished (1999)
COMMENT
Contact: Koehler K
MIPS
Am Klopferplatz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
clone from S. Wiemann, Molecular genome analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by BMFZ (Biomedical Research Center at the Charite,
Berlin/Germany) within the cDNA sequencing consortium of the German
Genome Project.
No sl sequence available.
This clone (DKFZP586E2117) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
location/Qualifiers
source
1. 757
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZP586E2117"
/clone_id="586 (synonym: hute1)"
/tissue_type="uterus"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSport1; site_1: NotI; site_2: SalI/MluI"

BASE COUNT 199 a 233 c 208 g 116 t 1 others

ORIGIN

alignment_scores:
Quality: 1129.00 Length: 258
Ratio: 4.552 Gaps: 8
Percent Similarity: 96.124 Percent Identity: 94.961

alignment_block:
us-09-668-119-3 x AL046886

```

```

Align seg 1/1 to: AL046886 from: 1 to: 757

1 MetArgLysAlaGlyValAlaHisSerLysSerLysAspMetGlnE 17
||||| 17
3 ArgGAAAGAGCTGTGTGTCACACAGTAATTCACAGCAAGATGAGAG 52
17 rHisValPheLeuLysAlaLysThrArgAspGluTyrLeuSerLeuVal 34
||||| 34
53 CCATGTTTCTCGAAGGCCAACACCCCGGACCAATACCTTCTCTCGGG 102
34 lArgLeuLeileHisPheArgAspIleHisAsnLysSerGlnAla 50
||||| 50
103 CCAGCTCATATTCATTTCCAGACATTCATTAACAGAAATCTCAACT 152
51 SerValSerAspProMetAsnAlaLeuGlnSerLeuThrGlyLysAl 67
||||| 67
153 TCCGTGATGATCTATGATGATGATGATGATGATGATGATGATGATG 202
67 aAlaGlyAlaAlaGlyIleGlyMetProProArgGlyProGlyGlnSer 84
||||| 84
203 TGGGGAGCGCGCTGGAATGGCATGCTCTCGGGGCCGAGAGTCTC 252
84 euGlyIleMetGlySerPheGlyAlaMetGlyGlnProMetSerLeuSer 100
||||| 100
253 TGGCGGGATGGGTAGCTTGTGTCATGGAGCAGCAGCAATGTCTCTCA 302
101 GlyGlnProProProGlyThrSerGlyMetAlaProHisSerMetAla 117
||||| 117
303 GGGCAGCGGCTCTCTGGACCTGGGGATGGCCCTCACAGCATGGCTCT 352
117 lValSerThrAlaThrProGlnThrGlnLeuGlnLeuGlnValAla 133
||||| 133
353 CGGTCTACGCGCAACTCCACAGCCAGCTGCAGCTCCAGAGGTGGCG 402
134 LeuGlnGlnGlnGlnGlnGlnGlnGlnPheGlnGlnGlnGlnAla 150
||||| 150
403 CTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCG 452
150 lAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnAla 166
||||| 166
453 CGGTACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCTCA 502
166 ngInSerAlaMetGlnGlnGlnPheGlnAlaValValGlnGlnGln 182
||||| 182
503 GCAGAGTGCATGCAGCAACAGTCCAAACAGTAGTGACAGCAGAGA 552
182 ngInLeuGlnGlnGlnGlnGlnGlnGlnGlnHisLeuLeuHis 199
||||| 199
553 GCAGCTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTCT 602
199 lsgInAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 215
||||| 215
603 ATCAAAATTCAGCAACAGATACAGCAGCAGCAGCAGCAGCTGCGGA 652
215 eAlaGlnLeuGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 232
||||| 232
653 AGCAGACGTGCAGCTC.....CAACAAACAGCAACAGCAGCAGCAG 696
232 lngGlnGlnGlnGlnAlaLeuGlnAlaGlnProProLiegGlnGlnPro 248
||||| 248
697 AGCAGCAGCAGCAGCAGCTTGTGAGGCCAGCCACCAATTCAGCAGCC 746
248 oMetGlnGln 251
||||| 251
747 GATGACAGCAG 756

seq_name: gb_est2:BG386117
seq_documentation_block:
LOCUS BG386117 878 bp mRNA EST 12-MAR-2001
DEFINITION 602455292F1 NIH_MGC_15 Homo sapiens cDNA clone IMAGE:4583495 5',
mRNA sequence.

```



Fri Mar 1 09:12:04 2002

us-09-668-119-3.rst

Page 6

ACCESSION BG386117  
VERSION BG386117.1 ST:13279563  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
1 (bases 1 to 878)  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
TITLE NIH-MGC http://mgi.mgi.nh.gov/  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-rti.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone distribution: NIH Intramural Sequencing Center  
Found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LCM1:07 row: h column: 24  
High quality sequence stop: 742.

FEATURES  
source Location/Qualifiers

1..878  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="4583495"  
/clone\_lib="NIH-MGC\_15"  
/rname\_type="adenocarcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: colon; Vector: pOT87; Site\_1: XhoI; Site\_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)"  
BASE COUNT 182 a 305 c 240 g 151 t  
ORIGIN

alignment\_scores:  
Quality: 1103.50 Length: 364  
Ratio: 4.294 Gaps:  
Percent Similarity: 70.604 Percent Identity: 67.857

alignment\_block:  
US-09-668-119-3 x BG386117 ..

Align seg 1/1 to: BG386117 from: 1 to: 878

27 AspGluTyrLeuSerLeuValAlaArgLeuIleIleHisPheArgAsp1 43  
|||||  
4 GACGAATACCTTCTCTCGTGGCAGGCTCATTCATTTCGAGACAT 53  
|||||  
43 eHisAsnLysLysSer31nAlaSerValSerAspPromeTasAlaLeuG 60  
|||||  
54 TCATACAGAGAATCTTAAGCTTCGCTGATCTATGATGACATCTC 103  
|||||  
60 InSerLeuThrGlyGlyProAlaAlaGlyAlaAlaGlyIleGlyMetPro 76  
|||||  
104 AAGAGCTGCTGGCGGACCTCTGCTGGGAGCGCTGGAATTTGGCAATGCTT 153  
|||||  
77 ProAlaGlyPProGlyGlnSerLeuGlyMetGlySerPheGlyAlaMe 93  
|||||  
154 CCTCGGGGCGCGGAGACATCTCTGGGCGGAGATGCTGAGCTTGGTCCAT 203  
|||||  
93 tGlyGlnPProMetSerLeuSerGlyGlnProProGlyThrSerGly 110  
|||||  
204 GGGACAGCCAAATGCTCTCTCAGGGCAGCCGCTCTGGAGCTGGGGA 253  
|||||  
110 eAlaProHisSerMeAlaValAlaSerThrAlaThrProGlnThrGln 126

|||||  
254 TGCCCTCCACAGCAGCAGCTGTCTGTCTACGCGCAATCTCCAGACCCAG 303  
|||||  
127 LeuGlnLeuGlnGlnValAlaLeuGlnGlnGlnGlnGlnGlnGln 143  
|||||  
304 CTGCAGCTCCAGCAGAGTGTGCTGCGCAGCAGCAGCAACAGCAGCAGTT 353  
|||||  
143 eGlnGlnGlnGlnGlnAlaAlaLeuGlnGlnGlnGlnGlnGlnGln 160  
|||||  
354 C..... 354  
160 InGlnGlnPheGlnAlaGlnGlnSerAlaMeGlnGlnPheGlnAla 176  
|||||  
177 ValValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 193  
|||||  
354 ..... 354  
193 sLeuIleLysLeuHisHisGlnAsnGlnGlnGlnGlnGlnGlnGln 210  
|||||  
354 ..... 354  
210 InGlnLeuGlnArgIleAlaGlnLeuGlnLeuGlnGlnGlnGln 226  
|||||  
354 ..... 354  
227 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 243  
|||||  
355 ..... 355  
243 oGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 260  
|||||  
390 AATTAGCAGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 439  
|||||  
260 eProGlnGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 276  
|||||  
440 TGCCCTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 489  
|||||  
277 ProGlnProGlnGlnProProValAlaGlnAsnGlnProSerGlnLeu 293  
|||||  
490 CCACAGCCCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 539  
|||||  
293 oProGlnSerGlnThrGlnProLeuValSerGlnAlaGlnAlaLeuPro 310  
|||||  
540 GCCACAGTCGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 588  
|||||  
310 LysGlnMetLeuTyrThrGlnProProLeuLysPheValAlaArgAlaProme 326  
|||||  
589 GACCAATGTG.TATACCCAAACCAACCACTGAAATTTTTCGAGCTCCGAT 637  
|||||  
326 tValValGlnGlnProProValGlnProGlnValAlaGlnGlnGln 342  
|||||  
638 GGTGTGTGAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 687  
|||||  
342 tAlaValAlaGlnThrAlaGlnAlaAlaGlnMetValAlaProGlyValGln 358  
|||||  
688 AGCAGTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 737  
|||||  
359 ValSerGlnSerSerLeu.....PromeLeuSerSerProse 371  
|||||  
738 ATGATTCACGGGGAGGAGCTTGGCCCGCAGGTGGAGATGCCATTAAGACCGG 787  
|||||  
371 tProGlyGlnGlnVal...GlnThrProGlnSer 381  
|||||  
788 TCCGCTTACCAAGGTGTGTGCTCCCTCCCGTCAAGC 821  
|||||  
seq\_name: gb\_est1:BE384305  
seq\_documentation\_block: 620 bp mRNA EST 21-JUL-2000  
LOCUS BE384305  
DEFINITION 601272684F1 NIH-MGC\_20 Homo sapiens cDNA clone IMAGE:3613868 5',  
mRNA sequence.



```

ACCESSION      BE384305
VERSION        BE384305.1 GI:9329670
KEYWORDS       EST.
SOURCE         Human.
ORGANISM       Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE      1 (bases 1 to 620)
AUTHORS       NIH-MGC http://mgc.ncl.nih.gov/
TITLE          National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL        Unpublished (1999)
COMMENT        Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-remail.nih.gov  
Tissue Procurement: ATCC/DCTD/DRP  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
Plate: LICM275 row: g column: 21  
High quality sequence stop: 620.
```

Location/Qualifiers

1..620
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3613868"
/clone_1lb="NIH MGC 20"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adapter: GGCAAGGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)." .

BASE COUNT      79 a            148 c            225 g            168 t

ORIGIN

alignment\_scores:

	Quality: 1059.00	Length: 206
	Ratio: 5.141	Gaps: 0
Percent Similarity: 100.000	Percent Identity: 99.515	

alignment\_block:

US-09-668-119-3 x BE384305/rev ..

Align seg 1/1 to reverse of: BE384305 from: 1 to: 620

172	G G A G C A T T C C A G C A C T A C T T C A G C A G C A G C A G C T C A C A G A G A	188
620	C A C A G C A G C A G C A G C A G C A G C A G C A G C A G C A G C A G A	571
188	n g i n g l i n g l i n h i s l e u i l e y l e u h i s h i s g l i n a s n g l i n g l i n t	205
570	G C A G C A G C A G C A G C A C A T C A A T T A A A T T G C A T C A T C A A A A T C A G A A C A G A	521
205	t e g l i n g l i n g l i n g l i n g l i n l e u g l i n a r y l l a l e g l i n e u g l i n e u g l i	221
520	T A C A G A G C A G C A C A G C A G C A C T C C A C C G A A T T G A C A C A G C T G C A G C T C C A A	471
222	G l i n g l i n g l i n g l i n g l i n g l i n g l i n g l i n g l i n g l i n g l i n a l a l e	238
470	C A C A G C A C A G C A G C A G C A G C A G C A G C A G C A G C A G C A G C T T T	421
238	i n g l i n a l a g l i n p r o t i l e g l i n a p r o p m e t g l i n i n p r o g l i n p r o p	255
420	G C A G G C C C A C C A C A A T T A G C A G C C A C G A T G C A G C A G C C A C A C C T T C	371
255	r o b r o s e r g l n a l a l e u p r o g l i n g l i n e u g l i n m e t h i s h i s t h r g i n	271

```

370 CGCCCTCCGAGGCTCTGCCCCAGACAGCTGCAGAGATGCATCAGACACAG 321
272 HISHISGInProProProGInInProGInGInInProProVaLaGInInaNGI 288
320 CACCACACAGCGCGCCACACAGCCCCAGCGAGCTCCAGTGGTCTCAGAACCA 271
288 nProSerGInLeuProProGInSerGInThrGInProLeuVaLSerGIna 305
270 ACCATCAACAACCTCCCGCCACAGTCGACAGACCAGCCTTGGTGTACAGG 221
305 IaGInaIaLeuProGlyGInMeLeuTyrThrGInProProLeuIysPhe 321
220 CGCAGGCTCTCCCTCGACAAATGTTATATACCCACACACACACTGAAATTT 171
322 ValATrgaIaProMeTValVaIaGInGInProProVaIaGInProGInVaIGI 338
170 GTCCGAGCTCCGATGCTGTGTGCAGACGCCCCAGTGCACACCCAGTGTCA 121
338 nGInGInGInThrIaVaIGInThrIaGInaIaIaGInMeTValaIaP 355
120 GCAGCAGCAGACAGCAGTACAGACACTCAGGCTGCCGAGTGGTGGCTC 71
355 roGlyVaIGInVaISerGInSerSerLeuProMeIeIeIeIeIeIeIeIeIe 371
70 CCGGAGTCCAGGTCCAGCAGAGAGCCTCCCAATGCTGTCTCGCCGCTCA 21
372 ProGlyGInGInVaIGIn 377
20 CCGGCGCAGCAGGTGCAG 3
seq_name: gb_est2:BG336790

```

```

seq_documentation_block:
LOCUS       BG336790          939 bp          mRNA          EST
DEFINITION  60240553F1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4542993 5',
            mRNA sequence.
ACCESSION   BG336790
VERSION     BG336790.1  GI:13143228
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 939)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT     Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgabbs@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.lnl.gov
            Plate: L1CML1224 row: a column: 10
            High quality sequence stop: 715.
            Location/Qualifiers
                1..939
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4542993"
                /clone_lib="NIH_MGC_21"
                /tissue_type="Choriocarcinoma"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: Placenta; Vector: pOTB7; Site_1: XhoI;
                Site_2: EcoRI; cDNA made by oligo-dT priming.
                Directionally cloned into EcoRI/XhoI sites using the
                following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                for average insert size 1.8kb. Library constructed by
                Ling Hong in the laboratory of Gerald M. Rubin (University
                of California, Berkeley) using ZAP-cDNA synthesis kit

```







Align seg 1/1 to: BG746380 from: 1 to: 914

```

402 AsnValSerSerGlyProAlaProSerProSerSerPheLeuProSerPr 418
|||||
4 AACGTGAGCTGTGGCCCTGGCCCATCTCCAGTAGCTTCTGCGCCAGCC 53
|||||
418 oSerProGlnProSerGlnSerProValThrAlaArgThrProGlnAsp 435
|||||
54 CTCACGCGAGCCCTCCAGAGCCAGTACGGCGGACCCCAACAACACT 103
|||||
435 heSerValProSerProGlyProLeuAsnThrProValAsnProSerSer 451
|||||
104 TCAGTGTCCCTTACTCTGAGACCTTTAAACACACCTTGTAACCCACTCT 153
|||||
452 ValMetSerProAlaGlySerSerGlnAlaGluGluGlnGlnTyrLeuAs 468
|||||
154 GTCATGAGCCGACAGCTGCTCCAGCCAGGCTGAGAGACACAGTACTGGA 203
|||||
468 ptySLeuGlyGlnLeuSerTyrTrpLeuProLeuArgArgMetIleA 485
|||||
204 CAAGCTGAAGCAGCTGTGCAAGTACATGAGCCCTGCGCCGATGATCA 253
|||||
485 snTysIleAspTyrAsnGluAspArgTyrTyrLeuAspLeuSerTyrMetTys 501
|||||
254 ACAAGATCCAGAACAGAACAGAACAGAACAGAACAGTAAAGTAAAG 303
|||||
502 SerLeuLeuAspIleLeuThrAspProSerTyrArgCysProLeuTyrTh 518
|||||
304 AGCCTTCTGTGACATCTGACAGACCCCTGCAAGCGGTGTCCTGTAAG 353
|||||
518 rLeuGlnTyrCysGluIleAlaLeuGlnTyrLeuTyrAsnAspMetArg 534
|||||
354 CTGCAAAAGTGTGAGTGTGCTGAGAAACTCAAGAAATGACATGGCGG 403
|||||
535 CysProLeuProHisArgProArgCysHisArgProAsnSerSerThrTy 551
|||||
404 TGCCCACTCCCGCCACCGCCCGCGTGCACCGCAACAACAGAGTACCTA 453
|||||
551 rAlaSerArgSerTyrMetProSerTyrProThrSerAlaHisLeuSerS 568
|||||
454 TGCAGCGCGCTCTGTGATGCTGCTCTGCGCCACATCCGCTCACCTGTCTT 503
|||||
568 erThrIleProCysThrAlaHisSerPheGlnPro 579
|||||
504 CAACCATTCCTGTACCGCAGCATTCCTTCCAGCCA 538

```

seq\_name: gb\_est2:BG746380

seq\_documentation\_block: 756 bp mRNA

LOCUS BG746380 EST 15-MAY-2001

DEFINITION 602703648F1 NIH\_MGC\_15 Homo sapiens cDNA clone IMAGE:4857083 5',

mRNA sequence.

ACCESSION BG746380

VERSION BG746380

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE NIH-MGC http://mgs.ncl.nih.gov/

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgaabs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

DNA Sequencing by: NIH Intramural Sequencing Center

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM1708 row: p column: 12

High quality sequence stop: 756.

location/Qualifiers

FEATURES

source

1..756

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4857083"

/clone\_1ib="NIH\_MGC\_15"

/issue\_type="adenocarcinoma cell line"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: colon; Vector: pORF7; Site\_1: XhoI; Site\_2:

EcoRI; cDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adapter: GGACGAGG(G). Size-selected >500bp for Hong in

insert size 1.8kb. Library constructed by Ling Hong in

the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)"

1 others

BASE COUNT 179 a 264 c 186 g 126 t

ORIGIN

alignment\_scores:

Quality: 919.00 Length: 176

Ratio: 5.251 Gaps: 1

Percent Similarity: 99.432 Percent Identity: 98.864

alignment\_block:

us-09-668-119-3 x BG746380 ..

Align seg 1/1 to: BG746380 from: 1 to: 756

```

405 SerGlyProAlaProSerProSerSerPheLeuProSerProSerProG 421
|||||
8 GCAGGCCAGCCCGCCATCTCCAGTAGCTTCTGCGCCACCCCTCAACGCA 57
|||||
421 nProSerGlnSerProValThrAlaArgThrProGlnAsnProSerValP 438
|||||
58 GCCCTCCAGAGCCCGAGTACGAGCGCGGAGCCCAACAAGTCAAGTGTCC 107
|||||
438 roSerProGlyProLeuAsnThrProValAsnProSerSerValMetSer 454
|||||
108 CCTCACCTGTGACCTTTAAACACACCTGTGAACCCCGCTGTATGATGAC 157
|||||
455 ProAlaGlySerSerGlnAlaGluGlnGlnGlnTyrLeuAspTyrLeu 471
|||||
158 CCAGCTGTGCTCCAGCCAGGCTGAGGAGCAGCAGTACCTGAGCAAGCTGA 207
|||||
471 sGlnLeuSerTyrTyrIleGluProLeuArgArgMetIleAsnTyrIleA 488
|||||
208 GCAGCTGTGAGTACATGACGACCCCTGCGCGGATGATCAACAAGATCG 257
|||||
488 sPlyAsnGluAspArgTyrTyrAspLeuSerTyrMetTyrSerLeuLeu 504
|||||
258 ACNAGAGACGAAAGCAAAAAAGGACCTGAGTAAAGATGAGAGAGCTTCTG 307
|||||
505 AspIleLeuThrAspProSerTyrArgCysProLeuTyrThrLeuGln 521
|||||
308 GACATTTGTACAGACCCCTGCAAGCGGTGTCCCTGGAAGACCTTGCACAA 357
|||||
521 sCysGluIleAlaLeuGluTyrLeuTyrAsnAspMetArgCysProLeu 537
|||||
358 GGTGTGAGATGCGCCCTGAGAAATCAAGAAATGATGAGGCGGCCACTTC 407
|||||
538 ProHisArgProArgCysHisArgProAsnSerSerThrTyrAlaSerAr 554
|||||
408 CCCACCGCGCCCGGTCGACCGACCAACAACAGCACTACTATGCGAGCGG 457
|||||
554 gSerThrMetProSerTyrProThrSerAlaHisLeuSerSerThrIleP 571
|||||
458 CTCTCGGATGCGGTCTGTGCGCAACATCCGCTACCTGCTTTCACCAATTC 507
|||||
571 roCysThrAlaHisSerPheGlnPro 579

```



Fri Mar 1 09:12:04 2002

us-09-668-119-3.rst

Page 10

508 CCCTGACCGCACATTTGTCCAGCCA 533

seq\_name: gb\_est1:BE019518

```
seq_documentation_block:
LOCUS      BE019568
```

LOCUS	526 bp	mRNA	EST	06-JUN-2000
DEFINITION	ba83f05.y1 N.H.MGC.21 Homo sapiens cDNA clone IMAGE:2907009 5 similar to TR:015413 O15413 CTG7A ; , mRNA sequence.			
ACCESSION	BE019568			

REVISION	27010000	
VERSION	BE019568.1	(I:8279646
KEYWORDS	EST.	
SOURCE	human	

ORGANISM

## REFERENCE

## REFERENCE AUTHORS

FILE  
JOURNAL

COMMENT

—

















Fri Mar 1 09:12:04 2002

us-09-668-119-3.rst

Page 14



Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result	Score	Query	Match	Length	DB	ID	Description
No.							
1	174.0	100.0	334	9	AF056191	AF056191 Homo sapi	
2	1703.8	97.9	3170	9	AF328769	AF328769 Homo sapi	
3	1070.4	61.5	3295	10	AF328770	AF328770 Mus muscu	
4	1035.8	59.5	1048	9	HS080745	U80745 Homo sapien	
5	659	37.9	2106	9	BC007529	BC007529 Homo sapi	
6	560	32.2	156315	2	AC068137	AC068137 Homo sapi	
7	353.4	20.3	64221	2	AL359452	AL359452 Homo sapi	
8	353.4	20.3	157080	2	AL358856	AL358856 Homo sapi	
9	353.4	20.3	163908	2	AC007050	AC007050 Homo sapi	
10	340.4	19.6	145356	5	AC004033	AC004033 Homo sapi	
11	338.6	19.5	386	6	AX071270	AX071270 Sequence	
12	220.8	12.7	238442	2	AC090437	AC090437 Mus muscu	
13	217.8	12.5	4701	2	AK000003	AK000003 Homo sapi	
14	209.8	12.1	121330	2	AC091616	AC091616 Rattus no	
15	209.6	12.0	187315	2	AC087802	AC087802 Mus muscu	
16	209.6	12.0	228467	2	AC079044	AC079044 Mus muscu	
17	207.2	11.9	5719	10	MMCA51	AF051726 Mus muscu	
18	206.8	11.9	162865	9	AC090645	AC090645 Homo sapi	
19	201.2	11.6	166874	33	AC021022	AC021022 Homo sapi	
20	201.2	11.6	168778	9	AC090886	AC090886 Homo sapi	
21	201.2	11.6	169590	9	AC090004	AC090004 Homo sapi	
22	201.2	11.6	209973	2	AC073755	AC073755 Mus muscu	
23	198	11.4	2479	10	RNM07960	U87960 Rattus norvegicus	
24	197.8	11.4	3941	10	RNLA02	X83546 R. norvegicus	
25	195.8	11.3	198470	2	AC090121	AC090121 Mus muscu	
26	191.4	11.0	197321	2	AC092203	AC092203 Mus muscu	
27	188.8	10.9	179533	2	AC024233	AC024233 Homo sapi	
28	185.4	10.7	197232	2	AC073946	AC073946 Mus muscu	
29	183.8	10.6	1611	10	MM070651	U70651 Mus musculus	
30	183.6	10.6	1612	10	MM070652	U70652 Mus musculus	
31	182.2	10.5	9553	5	AF134331	AF134321 Disosfic	
32	181.8	10.4	335884	2	AC092202	AC092202 Mus muscu	
33	181.6	10.4	1586	10	MM070053	U70053 Mus musculus	
34	181.6	10.4	1598	10	MM070054	U70054 Mus musculus	
35	181.2	10.4	3489	6	AR072952	AR072952 Sequence	
36	181.2	10.4	3489	6	AX107938	AX107938 Sequence	
37	181.2	10.4	3489	14	KSU52064	U52064 Kaposi's sar	
38	181.2	10.4	32207	6	AR065852	AR065852 Sequence	
39	181.2	10.4	32207	6	AR127850	AR127850 Sequence	
40	181.2	10.4	137508	14	KSU75598	U75598 Kaposi's sar	
41	180.8	10.4	28559	14	AF148805	AF148805 Kaposi's	
42	180.6	10.4	186935	2	AC022332	AC022322 Homo sapi	
43	177.4	10.2	180668	2	AC020857	AC020857 Mus muscu	
44	176.4	10.1	2070	5	BS043200	U43200 Borreogadus	
45	174.4	10.0	1968	10	MM046463	U46463 Mus musculus	

## ALIGNMENTS

RESULT	1
AF056191	
LOCUS	AF056191 3334 bp mRNA
DEFINITION	Homo sapiens TPA inducible protein mRNA, complete cds.
ACCESSION	AF056191
VERSION	AF056191.1 GI:3037134
KEYWORDS	.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Eukaryota; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 3334)
TITLE	Abraham,S. and Solomon,W.B.
	A novel glutamine-rich putative transcriptional adaptor protein
	(TIG-1), preferentially expressed in placental and bone-marrow
	tissues
JOURNAL	Gene 255 (2), 389-400 (2000)
PUBMED	11024300
REFERENCE	2 (bases 1 to 3334)
AUTHORS	Abraham,S. and Solomon,W.B.



Fri Mar 1 09:11:57 2002

us-09-668-119-1.rge

Page 2

TITLE Direct Submission  
JOURNAL Submitted (30-MAR-1998) Microbiology, State University of New York  
Health Science Center at Brooklyn, 450 Clarkson Ave., Brooklyn, NY  
11203, USA

FEATURES  
Source Location/Qualifiers

CDS  
1..3334  
/or\_janism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="K562"  
165..1904  
/note="transcription factor; rich in glutamine residues;  
TPA inducible gene; TIG-1"  
/colony\_start=1  
/product="TPA inducible protein"  
/protein\_id="AAC12944.1"  
/db\_xref="GI:3037135"  
/translation="MRKAGVHSHKSKDMESHVFLKAKTRDEYLSIVARLIHFPHDI  
NKKIQAASVSPMNAIOLSLGGPAAGAGIAMPBPGGSLGSGMSTGAMGPMISGCO  
PPTSTGMAPHSMAYVSTAPPTOTLOIOLVALIOOVALIOOVALIOOVALIOOVALIOO  
OORIAOOSAMCOOFOAVOOCIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOO  
OLOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOO  
POVIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOO  
POVIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOO  
GQVIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOO  
GQVIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOOIOO  
DMSKRPGLTKCEQALDKLQSLKSLIEFLKRMINKIDNEKRDLSKMSLDILT  
SSTLPETAHSPOR"

BASE COUNT 800 a 1088 c 893 g 553 t  
ORIGIN

Query Match 100.0%; Score 1740; DB 9; Length 3334;  
Best Local Similarity 100.0%; Pred. No. 5.4e-256;  
Matches 1740; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 atggagaaagctggtgtgagcaacaatgaatccagcaagataatgagagcatgtttc 60  
165 ATGAGGAAAGCTGGTGTGAGCAACAGTAATCCAGCAAGATATGAGAGCATGTTTC 224  
61 ctgaaagcaagacccgagagcaatacttctctgagcagagcatatcatcattt 120  
225 CTGAAGGCCAAGACCCGAGAGCAATACCTTCTCTGTGGCCAGGCTCATATCCATT 284  
121 cgagacatcataaagaagaatctcaagcttcgctgaagatctgaatgaatcag 180  
285 CGAGACATTCATAAAGAAATCTCAAGCTTCCTGATGATCTATGATGACATCCAG 344  
181 agcctgactgagcctctctgagcagcagcgtgaaattgacatgctcctcccgaggccg 240  
345 AGCCTGACTGAGCCTCTCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 404  
241 ggaagctctctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 300  
405 GGACAGCTCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 464  
301 ggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 360  
465 GGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 524  
361 ggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 420  
525 GGAAGCTCCAGAGCCAGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 584  
421 cagcagctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 480  
585 CAGCAGCTCCAGAGCCAGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 644  
481 cagcagctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 540  
645 CAGCAGCTCCAGAGCCAGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 704  
541 cag 600

705 CAGCAGAGCTCCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 764  
601 aatcagcaacagatacagcagcagcaacagcagcagcagcagcagcagcagcagcagcagc 660  
765 AATCAGCAACAGATACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 824  
661 caacaacagcaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 720  
825 CAACAACAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 884  
721 cagcagcaatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 780  
885 CAGCAGCAATTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 944  
781 cccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 840  
945 cccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1004  
841 cagcctcagctgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 900  
1005 CAGCCTCCAGCTGTCTCAAGAACCAACATCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1064  
901 ttggtgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 960  
1065 TTGGTGTCAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1124  
961 ttgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1020  
1125 TTGTGTCCAGCTCCAGCTGTCTCAAGAACCAACATCCAGCAGCAGCAGCAGCAGCAGCAG 1184  
1021 cagc 1080  
1185 CAG 1244  
1081 cagc 1140  
1245 CAG 1304  
1141 ttggtgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1200  
1305 TCGATGCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 1364  
1201 tccaaagctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1260  
1365 TCCAAAGCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1424  
1261 cagcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1320  
1425 CAGCCTCCAGAGCCAGCTGTCTCAAGAACCAACATCCAGCAGCAGCAGCAGCAGCAGCAG 1484  
1321 ggaagcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1380  
1485 GGACCTTTAAACACACACCTGTGAACCCAGCTGTGATGAGCCAGCTGAGCTCCAGCAG 1544  
1381 gctgagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1440  
1545 GCTGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1604  
1441 cgcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1500  
1605 CGCCGATGATCAACAGATGACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1664  
1501 aagagcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1560  
1665 AAGAGCCTTTGAGCAGCTTCTGACAGAGCCCTTCGAGAGCGGTCTCCCTGAGAGCCTTG 1724  
1561 aagtgtagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1620  
1725 AAGTGTGAGATCGCCCTGAGAGAACTCAAGAAATGACATGCGGTGCCACTCCACACCG 1784  
1621 cccgcgtagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1680  
1785 CCGCGGTGCCAGCAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1844











QY 583 attaatgcatcatcaaatcaatcagcaagatatacagcagcagcaacagcagctgcagcga 642  
 Db 673 ATTAAATTGATCATATCAAGCCAGCAACA--ACAGATACACACACAGCACTGCAGAG 729  
 QY 643 atagacagctgcagctccaacaacagcaacagcagcagcagcagcagcagcagcagcag 702  
 Db 730 ATGGCAGATTGTCAGCT-----GCAACAACACCAACAGCAGCAACAG 771  
 QY 703 cagcagcgtttgagagccagcagcccaatcaagcaacagcagcagcagcagcagcagcct 762  
 Db 772 CAGCAGAGCTTTTCAGAGCCAGCCAGCAATGCGAGCCATCATCATCCACAGCAGCT 831  
 QY 763 cagcagcgttcagcagctgcagcagcagcagcagcagcagcagcagcagcagcagcag 822  
 Db 832 CCCCCCTTCTCAGAGCCCTTACCCAGCAGCTGTACAGCTGATCATCCACAGCATCAGCA 891  
 QY 823 ccgcaacacagcagccagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 882  
 Db 892 CCACCACTCAGAGCTCAGAGAGTCCCCCATTTGCTCAAAAACCAACACAGATGCCACCA 951  
 QY 883 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 941  
 Db 952 CAGTACACAGCAGCAGCTTTGGTGTACAGCAGCAAGCCCTTCCAGCCGATCCTGTAT 1011  
 QY 942 --taaccacacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 999  
 Db 1012 GCTGCCCAACAGCAGCTGAAATTTTCTGCTCCGATGCTGTGTCAGCAGCCGCAAGTG 1071  
 QY 1000 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1041  
 Db 1072 CAGCCCAAGCTCAGAGAGTGCAGCCCAAGCTGTCAGAGCCGCAAGCTGTCAGAGCAGCA 1131  
 QY 1042 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1071  
 Db 1132 CAGTGTGCCAGATGATGCTCCGAGCTCAGATGATGTCGAAGCCTTGCCCAAGGC 1191  
 QY 1072 ----- 1071  
 Db 1192 GGGATCAGGTAAAGAGCCGGTTCCGCCACCTCCACATGTCGTGCGCCGTCAGAC 1251  
 QY 1072 -----cagtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1101  
 Db 1252 TCCATCTTTGGGCGGAGCCACACAGAGCTCAGCAAGCAAGCCTCAGCATGCTG 1311  
 QY 1102 tcttcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1161  
 Db 1312 TCTTACCGCTCAGCAGGCGAGGTCAGACCCACAGTCGATGCTCCGCCACAG 1371  
 QY 1162 ccgtcccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1221  
 Db 1372 CGGTCCCAACACTG-----GTCACAGCCCAACTCTAATGTACGTCCGCGCT 1422  
 QY 1222 gcccaatctccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1281  
 Db 1423 GCCCATCTCCAGCAGCTCTGCTAGCCCTTACACACAGCTCTCAGAGCCGCGTG 1482  
 QY 1282 aagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1341  
 Db 1483 ACAGCAGGAGCCCAAGACCTTACGCTTCTCCCTGAGCTTAAACACCCCTGTG 1542  
 QY 1342 aacccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1401  
 Db 1543 AACCCAGCTGTATGAGCCAGCTGTAGCCAGCTGAGAGAGCAGCAGTATCCTG 1602  
 QY 1402 gacaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1461  
 Db 1603 GACAACTGAAGCAATTTGCCAAGTACATGAGCCCTCGAGCATGATCAACAAGATC 1662  
 QY 1462 gacaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1521  
 Db 1663 GACAAGATTAAGCAAGAAAAAGACTTAAGTAAATGAGAGCCTGCTGAGCATCTC 1722  
 QY 1522 acagacccctcgaagcgtgtccctgaagaccttgcaaatgttgatgcgccccgag 1581

Db 1723 ACGACCCCTCGAAGAGGTGTCCTCCCTGAAGACCTGTGAAAAAGTGTGATGGCTGAGAG 1782  
 QY 1582 aaactcaagaatgacat-gcagtgccatcctcccccagcagcagcagcagcagcagcagc 1640  
 Db 1783 AAGCTCAAGATGACATGAGGAGTGGCCACACACCCACACACCCAGTCTTCCAAACCA 1842  
 QY 1641 cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1700  
 Db 1843 CAGCAGAGCCTGTCCCAACCACTCTGTAGTCAAGTCTTGGCCACATCCGTTACCTGTC 1902  
 QY 1701 ttcaacatccctgtacccagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1740  
 Db 1903 TTCAACCATTCCTGTACCGCAGCATTTGTGCGAGCCATGA 1942

## RESULT 4

HSU80745

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

SOURCE

gene

CDS

repeat\_region

BASE COUNT

ORIGIN

Query Match

Best local Similarity

Matches 1037;

Conservative

Score 1035.8;

DB 9;

Length 1048;

Pred. No. 1,2e-148;

Mismatches 3;

Indels 0;

Gaps 0;

1037;

Conservative

Score 1035.8;

DB 9;

Length 1048;

Pred. No. 1,2e-148;

Mismatches 3;

Indels 0;

Gaps 0;

1037;

Conservative











```

OY 765 ggcctccagagctgccccagagctgacagatgacatcaacacagaccacagcc 824
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 154274 cccctccagagctgccccagagctgacagatgacatcaacacagaccacag 154215
OY 825 gccacacagccccagagctgacagctgacagatgacatcaacacagaccacag 882
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 154214 gctggcagatgctccagacacacacacacacacacacacacacacacacac 154155
OY 883 --caatgcagacacacacacacacacacacacacacacacacacacacacacac 940
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 154154 gctgacacacacacacacacacacacacacacacacacacacacacacac 154098
OY 941 ataccacacacacacacacacacacacacacacacacacacacacacacacac 1000
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 154097 atgccacacacacacacacacacacacacacacacacacacacacacacac 154038
OY 1001 agccccaagt-----gcagcagcagcagcagcagcagcagcagcagcagc 1043
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 154037 agccccaagt-----gcagcagcagcagcagcagcagcagcagcagcagc 153978
OY 1044 ggcctccagagctgccccagagctgacagatgacatcaacacagaccacag 1103
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 153977 ggcctccagagctgccccagagctgacagatgacatcaacacagaccacag 153918
OY 1104 ctgcgcagacacacacacacacacacacacacacacacacacacacacacac 1163
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 153917 ctgcgcagacacacacacacacacacacacacacacacacacacacacacac 153858
OY 1164 ctgcgcagacacacacacacacacacacacacacacacacacacacacacac 1221
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 153857 gctgcacacacacacacacacacacacacacacacacacacacacacacac 153798
OY 1222 gccccaatgccag-----taactctgcccagccctccacacacacacacac 1277
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 153797 gccccaatgccag-----taactctgcccagccctccacacacacacacac 153738
OY 1278 agtgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1337
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 153737 agtgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 153684
OY 1338 tctgaacacacacacacacacacacacacacacacacacacacacacacacac 1397
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 153683 tctgaacacacacacacacacacacacacacacacacacacacacacacacac 153624
OY 1398 cctgcagacacacacacacacacacacacacacacacacacacacacacacac 1457
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 153623 cctgcagacacacacacacacacacacacacacacacacacacacacacacac 153564
OY 1458 gctgcagacacacacacacacacacacacacacacacacacacacacacacac 1517
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 153563 gctgcagacacacacacacacacacacacacacacacacacacacacacacac 153508
OY 1518 tctgaacacacacacacacacacacacacacacacacacacacacacacacac 1577
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 153507 tctgaacacacacacacacacacacacacacacacacacacacacacacacac 153448
OY 1578 ggaagaacacacacacacacacacacacacacacacacacacacacacacacac 1636
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 153447 ggaagaacacacacacacacacacacacacacacacacacacacacacacacac 153388
OY 1637 caaagaacacacacacacacacacacacacacacacacacacacacacacacac 1693
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 153387 caaagaacacacacacacacacacacacacacacacacacacacacacacacac 153328
OY 1694 accgtctcaacacacacacacacacacacacacacacacacacacacacacacac 1737
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 153327 -gcagtctcgaacacacacacacacacacacacacacacacacacacacacacacac 153285

```

```

RESULT 7
AL359452 6..231 bp DNA HTG 13-JUN-2001
LOCUS

```

```

DEFINITION Homo sapiens chromosome 1 clone RP11-422P18, *** SEQUENCING IN
ACCESSION PROGRAMS ***, 20 unordered pieces.
VERSION AL359452
KEYWORDS GI:9930964
SOURCE HTGS-PHASE1; HTGS-CANCELLED.
ORGANISM human.
REFERENCE Homo sapiens
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL 1 (bases 1 to 64231)
Direct Submission
Submitted (12-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquey@sanger.ac.uk; Clome
requests: clomequest@sanger.ac.uk
On Aug 27, 2000 this sequence version replaced gi:9864141.
COMMENT
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquey@sanger.ac.uk
----- Project Information
Center project name: BA422P18
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 52454 bases at least Q40
Consensus quality: 57067 bases at least Q30
Consensus quality: 59753 bases at least Q20
Insert size: 62331; sum-of-contrigs
Insert size: 190157; 8.2% error; agarose-fp
Quality coverage: 2.09x in Q20 bases; sum-of-contrigs Quality
coverage: 1.09x in Q20 bases; agarose-fp

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 20 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved. 2992: contrig of 2992 bp in length
* 2993 3092: gap of 100 bp
* 3093 5136: contrig of 2044 bp in length
* 5137 5236: gap of 100 bp
* 5237 7850: contrig of 2614 bp in length
* 7851 7950: gap of 100 bp
* 7951 9976: contrig of 2026 bp in length
* 9977 10076: gap of 100 bp
* 10077 12531: contrig of 2455 bp in length
* 12532 12631: gap of 100 bp
* 12632 14688: contrig of 2057 bp in length
* 14689 14788: gap of 100 bp
* 14789 18056: contrig of 3268 bp in length
* 18057 18156: gap of 100 bp
* 18157 21784: contrig of 3628 bp in length
* 21785 21884: gap of 100 bp
* 21885 24801: contrig of 2917 bp in length
* 24802 24901: gap of 100 bp
* 24902 27603: contrig of 2702 bp in length
* 27604 27703: gap of 100 bp
* 27704 30911: contrig of 3208 bp in length
* 30912 31011: gap of 100 bp
* 31012 35721: contrig of 4710 bp in length
* 35722 35821: gap of 100 bp
* 35822 38454: contrig of 2633 bp in length
* 38455 38554: gap of 100 bp
* 38555 41225: contrig of 2671 bp in length
* 41226 41325: gap of 100 bp
* 41326 48066: contrig of 6741 bp in length
* 48067 48166: gap of 100 bp

```































Fri Mar 1 09:11:57 2002

us-09-668-119-1.rge

Page 16

Db 72241 CAGAGCCAGCCTTGTGTGTCACAAGCACAAGCCCTTCTGTGACCGATGCTGTATGCTGCC 72182

Qy 946 caaccacactgaatttgt 965  
|||||  
|||||

Db 72181 CAGAGCAGCTGAAATTGT 72162

Search completed: February 28, 2002, 19:26:56  
Job time: 10749 sec